

DIGITHÈQUE

Université libre de Bruxelles

J.-J. Droesbeke, B. Fichet, Ph. Tassi, éd., *Modélisation ARCH*, Editions de l'Université de Bruxelles, 1994.

http://digistore.bib.ulb.ac.be/2007/DL2826626_000_f.pdf

Cette œuvre littéraire est soumise à la législation belge en matière de droit d'auteur.

Elle a été publiée par les
Editions de l'Université de Bruxelles
<http://www.editions-universite-bruxelles.be/>

Les règles d'utilisation de la présente copie numérique de cette œuvre sont visibles sur la dernière page de ce document.

L'ensemble des documents numérisés mis à disposition par les bibliothèques de l'ULB sont accessibles à partir du site
<http://digitheque.ulb.ac.be/>

ASSOCIATION POUR LA STATISTIQUE ET SES UTILISATIONS

MODÉLISATION ARCH

THÉORIE STATISTIQUE ET
APPLICATIONS DANS LE DOMAINE
DE LA FINANCE

ÉDITÉ PAR JEAN-JACQUES DROESBEKE,
BERNARD FICHET, PHILIPPE TASSI



ÉDITIONS
DE L'UNIVERSITÉ
DE BRUXELLES

ASU

ÉDITIONS
ELLIPSES



MODÉLISATION ARCH

**THÉORIE STATISTIQUE ET
APPLICATIONS DANS LE DOMAINE
DE LA FINANCE**

MODÉLISATION ARCH

**THÉORIE STATISTIQUE ET
APPLICATIONS DANS LE DOMAINE
DE LA FINANCE**

**ÉDITÉ PAR JEAN-JACQUES DROESBEKE,
BERNARD FICHET, PHILIPPE TASSI**

**Des mêmes éditeurs scientifiques,
chez Economica**

Les sondages, 1987.

Séries chronologiques: théorie et pratique
des modèles ARIMA, 1988.

Analyse statistique des durées de vie, 1988.

Modèles pour l'analyse des données
multidimensionnelles, 1992.

© 1994 by Editions de l'Université de Bruxelles
Avenue Paul Héger 26 - 1050 Bruxelles (Belgique)
D/1994/0171/1
ISBN 2-8004-1081-7

Ellipses
Editions Marketing
32 rue Bargue - 75015 Paris (France)
ISBN 2-7298-9422-5

Imprimé en Belgique

Un savetier chantait du matin jusqu'au soir :
C'était merveilles de le voir,
Merveilles de l'ouïr ; il faisait des passages,
Plus content qu'aucun des sept sages.
Son voisin au contraire, étant tout cousu d'or,
Chantait peu, dormait moins encor.
C'était un homme de finance.

Jean de la Fontaine, *Fables*, VIII, II.
Le savetier et le financier, v. 1-7.

AVANT-PROPOS

Les *Journées d'Etude en Statistique* (J.E.S.) ont été organisées pour la cinquième fois, en 1992, par l'*Association pour la Statistique et ses Utilisations* (A.S.U.) avec le concours de la Société Mathématique de France.

Les premières journées, réalisées en octobre 1984, ont été consacrées à l'*Analyse des séries chronologiques*, dans la lignée de la méthodologie de Box et Jenkins (problèmes liés à la spécification, à l'estimation et à la validation de modèles stochastiques). En octobre 1986, se sont déroulées les deuxièmes J.E.S. avec pour thème les *Sondages*. Notre choix avait été guidé par l'intérêt né des développements en la matière, tant en théorie qu'en pratique. En octobre 1988, les J.E.S. ont abordé les problèmes de l'analyse statistique des *durées de vie*. La structure de ces journées avait pour but de montrer que les concepts et les méthodologies liées à ce sujet présentent un intérêt général et méritent d'être vus sous un aspect statistique et pluraliste. En octobre 1990, les J.E.S. ont porté sur les *modèles pour l'analyse des données multidimensionnelles*, centrant leur attention sur les rapports pouvant exister entre l'approche probabiliste «classique» et l'analyse des données à la «française».

Du 19 au 23 octobre 1992 se sont déroulées les cinquièmes J.E.S. Elles étaient consacrées aux *Modèles ARCH et à leurs applications à la finance*. Dans la lignée de l'utilisation des processus stochastiques de type ARMA pour le traitement des séries chronologiques, cette classe de modèles non linéaires permet une étude complète, illustrée par des applications dans des domaines concernant principalement des problèmes financiers et monétaires.

Nous remercions les divers conférenciers qui ont prêté leur concours à ces Journées : Jean-François BOULIER (Crédit Commercial de France), Laure ELIE (Université Paris VII), Christian GOURIEROUX (C.R.E.S.T.), Guy MELARD (Université Libre de Bruxelles), Alain MONFORT (INSEE) et Eric RENAULT (Université des Sciences Sociales, Toulouse I).

Nous tenons aussi à remercier particulièrement Thérèse Lekeux-Sonck qui a bien voulu gérer la composition de cet ouvrage, ainsi que toutes les personnes qui, à titre divers, nous ont apporté leur aide, que ce soit au Centre International de Rencontres Mathématiques de Luminy, à l'Université Libre de Bruxelles, au Laboratoire de Biomathématiques de l'Université d'Aix-Marseille II ou à Médiamétrie.

Jean-Jacques Droesbeke
Université Libre de Bruxelles

Bernard Fichet
Université d'Aix-Marseille II

Philippe Tassi
Médiamétrie

CHAPITRE 1

LES MODÈLES ARCH ET LEUR HISTOIRE

par

Jean-Jacques DROESBEKE,
Bernard FICHET, Philippe TASSI

La théorie financière recourt de façon significative aux outils statistiques depuis plus de trente ans. On peut ainsi mentionner le travail tout à fait remarquable réalisé par Louis Bachelier (1870-1946), qui, dans sa thèse de doctorat es Sciences mathématiques, défendue en mars 1900 et intitulée « *Théorie de la Spéculation* », introduisit le concept de marché efficient bien longtemps avant que cette notion soit développée avec l'intérêt que l'on connaît. Il utilisa à cet effet des modèles de marche aléatoire, des mouvements browniens et des martingales. Il se posa même la question de tester sa théorie empiriquement. Mais son œuvre resta discrète jusqu'en 1960, date de la traduction anglaise de son travail.

Jusque dans les années 1950, les ouvrages consacrés à la finance furent très souvent descriptifs. Le but essentiel consistait à décrire et informer sur les instruments financiers, les institutions financières et, de façon générale, les pratiques financières des entreprises (cf. par exemple, Solomon [1967]). Parmi les travaux qui bouleversèrent cette situation, il faut notamment mentionner ceux de Markowitz [1952, 1959] et Tobin [1958] sur la sélection de portefeuilles d'une part, ceux de Modigliani et Miller [1958] sur la structure du capital et l'évaluation des firmes, d'autre part.

En ce qui concerne l'usage des méthodes statistiques en finance, plusieurs voies ont été suivies.

- a) L'utilisation des *modèles de régression* et des *modèles économétriques* se retrouve dans pratiquement tous les secteurs de l'analyse financière. La référence la plus marquante en la matière faisant l'état de la question pour les années 1960 à 1975 est certainement l'ouvrage écrit par Fama en 1976 et intitulé *Foundations of finance*. Cette publication ne doit cependant pas nous faire oublier, comme indiqué ci-dessus, les travaux basés sur les publications de Markowitz [1959] dans la sélection des portefeuilles, qui ont débouché sur des modèles de marché relativement complexes et qui ont constitué un domaine de réflexion théorique intéressant pour la recherche statistique.
- b) *L'analyse multivariée* (analyse en composantes principales, analyse discriminante, ...) a aussi constitué un outil de plus en plus utilisé dans de nombreuses études exploratoires. Citons en particulier les articles de Pinches et Mingo [1973] et ceux de Herbst [1974] qui recourent à la fois à une analyse factorielle et une modélisation par régression multiple, sans oublier les travaux de Lloyd et Lee [1976] sur des modèles d'équilibre des actifs financiers. L'analyse discriminante a aussi trouvé dans l'usage des modèles logit une alternative utilisée par les praticiens.
- c) L'usage des *modèles de séries chronologiques*, et tout particulièrement ceux associés à la classe des processus ARMA, constitue un aspect important de l'application de la statistique en finance. Dès le milieu des années 1970, plusieurs auteurs ont recouru à la méthodologie de Box et Jenkins pour estimer ou prévoir gains et taux. Nous évoquerons l'usage de cette démarche ci-dessous. Il est cependant utile de mentionner aussi dans cette catégorie, les tentatives liées à l'usage de l'analyse spectrale, notamment appliquée à l'étude des taux d'intérêt ou pour tester l'efficacité d'un marché (cf., par exemple, Granger et Morgenstern [1970] et Percival [1975]).
- d) *La théorie de la décision* constitue un quatrième outil, permettant à divers auteurs de recourir à une approche bayésienne dans leur démarche : citons en particulier les travaux de Winkler et Barry [1975] dans le choix d'un portefeuille et ceux de Vasicek dans l'estimateur des betas des actions [1973].

Le développement des modèles ARCH se place dans le contexte et la lignée des modèles de séries chronologiques évoqués ci-dessus. Ces modèles ont été essentiellement développés avec des objectifs de description, de désaisonnalisation, de prévision ou de contrôle de systèmes. L'âge d'or de cette modélisation se situe dans les années 1970 avec le développement des modèles autorégressifs-moyennes mobiles (ARMA) et de leurs généralisations, qui présentaient l'avantage de se prêter facilement à l'emploi. Comme nous l'avons déjà souligné plus haut, leur usage s'est trouvé facilité par le recours à une méthodologie, due à G.E.P. Box et G.M. Jenkins, destinée à aider l'utilisateur dans le choix d'un modèle, l'estimation de ses paramètres et sa validation, méthodologie qui depuis une vingtaine d'années a engendré des travaux aussi multiples qu'intéressants.

Ces modèles sont fondés sur une expression de la valeur présente d'une variable comme fonction linéaire de ses valeurs passées et de valeurs présentes et passées d'un bruit aléatoire, s'interprétant comme l'innovation de la série.

Plus précisément, on appelle processus autorégressif moyenne mobile d'ordre (p, q) un processus stationnaire $(Y_t; t \in \mathbb{Z})$ satisfaisant à l'équation :

$$Y_t - \sum_{i=1}^p \phi_i Y_{t-i} = \theta_0 + \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} \tag{1.1}$$

où les ϕ_i et θ_i sont des paramètres réels et $(\varepsilon_t; t \in \mathbb{Z})$ est un bruit blanc de variance σ^2 (cf. chapitre 2). Cette formulation présente cependant quelques inconvénients. Elle est linéaire, ce qui restreint sa généralité. Elle est, en outre, mise en œuvre en imposant peu de contraintes sur les paramètres autorégressifs, ce qui empêche des interprétations plus structurelles des problèmes étudiés.

Parmi les domaines d'applications où la modélisation ARMA se révèle insuffisante, figurent certains problèmes financiers et monétaires. Les séries disponibles dans ce secteur présentent en effet souvent des caractéristiques de dynamique non linéaire, dont la plus significative est le fait que la variabilité instantanée de la série (appelée *volatilité*) dépend de façon importante du passé. Il existe d'autre part des théories financières basées sur des principes d'équilibre et de comportements rationnels des agents intervenant sur le marché qui conduisent naturellement à introduire et à tester des contraintes structurelles sur les paramètres.

Si l'on se place d'un point de vue statistique, les modèles ARCH constituent une classe spécifique de modèles non linéaires pour laquelle on peut mener une étude complète en abordant un certain nombre de problèmes classiques : test de marche aléatoire, détermination d'intervalles de prévision, recherche de séries sous-jacentes (facteurs), ...

Historiquement, les modèles ARCH (*Autorégressifs Conditionnellement Hétéroscédastiques*) ont été introduits par R.F. Engle en 1982. Dans son article, l'auteur ne suppose plus que $(\varepsilon_t; t \in \mathbb{Z})$ est un bruit blanc mais envisage plutôt que ce processus est de la forme :

$$\varepsilon_t = v_t \sqrt{h_t} \tag{1.2}$$

où

$$v_t \sim \text{i.i.d. } N(0, 1) \tag{1.3}$$

$$h_t = c + \sum_{i=1}^q a_i \varepsilon_{t-i}^2. \tag{1.4}$$

Dans cette expression, on suppose que $c > 0$ et que $a_i \geq 0$ ($i = 1, \dots, q$). Cette façon de procéder permettait à Engle de tenir compte du fait que les variations de prix — fortes ou faibles — étaient suivies d'autres variations fortes ou faibles

de signes imprévisibles. Certaines conditions étaient en outre imposées afin de réduire le nombre de paramètres du modèle.

La lecture de la bibliographie présentée dans cet ouvrage souligne à quel point nous trouvons encore dans une phase expérimentale et en pleine évolution. D'ores et déjà, les modèles ARCH ont donné lieu à des extensions et utilisations diverses (modèles GARCH, ARMA à erreur ARCH, ...).

Par ailleurs, la théorie financière a été repensée de façon à inclure la dépendance temporelle de la volatilité. Dès lors, à côté des développements de la théorie des modèles ARCH, on assiste à une remise en question de l'étude concernant un certain nombre de problèmes, comme les déterminations de portefeuilles optimaux, les évaluations de prix des options, ...

En ce qui concerne les domaines d'application, on peut en distinguer deux grandes catégories. Les premiers consistent à tester des théories économiques relatives aux divers marchés (devises, obligations, ...), à se pencher sur les liens entre court et long termes. Les seconds traitent des comportements d'intervention sur le marché des établissements financiers (détermination de portefeuilles optimaux, de portefeuilles de couverture, ...). Ce dernier type d'application est plus « sensible » et, par là-même, généralement couvert par le « secret bancaire ».

Le cours dispensé dans le cadre des *Journées d'Etude en Statistique* d'octobre 1992 a été imprégné de cette situation, présentant à la fois une composante d'enseignement traditionnel et le souci d'exposer des résultats ou thèmes de recherche très récents.

La structure de cet ouvrage reprend ce principe. Le chapitre 2 est consacré à une présentation générale des modèles linéaires et non linéaires et de leurs caractéristiques par Guy MELARD. Constituant un chapitre de base, il permet au lecteur une (re)mise à niveau indispensable.

Le chapitre 3 est destiné à décrire les diverses modélisations que l'on peut classer sous la rubrique des modèles hétéroscédastiques univariés et certaines de leurs propriétés. Il est dû à Christian GOURIEROUX comme d'ailleurs le chapitre suivant consacré à des problèmes d'estimation (méthode du pseudo-maximum de vraisemblance — méthode en deux étapes et méthode par variables instrumentales), de prévision et de tests (hypothèse de marche aléatoire, homoscedasticité).

Le chapitre 5 a pour auteur Alain MONFORT. Il porte sur des généralisations et alternatives de la modélisation ARCH. On y trouve tout d'abord le cas des modèles multivariés dont l'étude est facilitée grâce à l'introduction de modèles dynamiques à facteurs. Le recours à des modèles à volatilité stochastique discrète et continue est également abordé. Certaines méthodes d'estimation adaptées à ces modèles sont considérées. Des méthodes plus sophistiquées sont présentées par le même auteur dans le chapitre 6. On y trouvera en particulier la méthode des moments simulés, la méthode du maximum de vraisemblance simulée, la méthode du pseudo-maximum de vraisemblance simulée et les méthodes d'inférence indirectes.

Le chapitre 7, dû à Sergio PASTERELLO et Eric RENAULT, a pour but d'étendre les interactions entre théories financière et statistique. Basés sur la notion de facteur en finance, différents modèles sont examinés. En outre, une attention toute particulière est portée sur l'évolution de performance de portefeuilles.

Laure ELIE consacre le chapitre 8 aux processus ARCH comme approximations de processus en temps continu dont l'utilisation dans le domaine de la finance est bien connue. De nombreuses raisons (simplicité des théories financières, rapidité des calculs, apparition de la cotation des titres en continu, ...) soulignent l'intérêt de cette matière.

Enfin, dans le dernier chapitre, Jean-François BOULIER s'interroge sur l'utilité des modèles ARCH. Cette réflexion se situe à la rencontre des théoriciens et praticiens et constitue sans contexte une invitation à approfondir ce domaine.

CHAPITRE 2

MODÈLES LINÉAIRES ET NON LINÉAIRES

par

Guy MÉLARD^(*)

2.1 INTRODUCTION

Dans ce chapitre, nous considérons des séries temporelles unidimensionnelles (c'est-à-dire relatives à une variable unique) ou multidimensionnelles dont les dates d'observation sont supposées équi-espacées, pour la simplicité. L'indice de temps peut alors être entier, entre 1 et T , où T est la longueur de la série. Une série $(Y_t, t = 1, \dots, T)$ peut être considérée comme une réalisation particulière d'un processus stochastique.

Depuis le début des années 1970, suite à la parution du livre de Box et Jenkins [1970; édition révisée en 1976], l'accent a été mis sur les *modèles linéaires*. Depuis une dizaine d'années, devant les problèmes de modélisation rencontrés pour certaines séries de données (les nombres de taches solaires, des séries de biologie, des séries financières et monétaires) les modèles non linéaires ont été envisagés. Plusieurs classes de modèles ont été abordées : les *modèles bilinéaires*, les *modèles à seuils*, les *modèles hétéroscédastiques*. L'étude de ces modèles n'est pas terminée mais a progressé de manière significative (Priestley [1988], Tong [1990], Gouriéroux [1992], Zakoian [1992]). Les modèles non

^(*) Je remercie Mme Baldassarini, ISTAT, Rome, et M. Charlier, Institut Economique Agricole, Bruxelles, qui m'ont aimablement fourni les données utilisées dans le paragraphe 2.3, ainsi que L. Broze, C. Gouriéroux et O. Scaillet pour leurs commentaires.

linéaires gagnent le domaine des applications (De Gooijer et Kumar [1992], Bollerslev *et al.* [1992]).

Après quelques rappels sur les processus stochastiques et la représentation de séries temporelles par des modèles linéaires, nous introduisons ces principales classes de modèles non linéaires.

2.2 SÉRIES TEMPORELLES ET PROCESSUS STOCHASTIQUES

Un *processus stochastique* ou *processus aléatoire* est une suite de variables aléatoires (définies sur le même espace probabilisé Ω) indicées par t . On utilise la même notation pour l'observation et la variable aléatoire, Y_t . Le processus Y est, selon le cas, $(Y_t, t \in \mathbb{Z})$ ou $(Y_t, t \in \mathbb{N})$ (\mathbb{Z} et \mathbb{N} sont respectivement l'ensemble des entiers et des naturels). Pour une première introduction à la représentation de séries temporelles par des processus stochastiques, voir par exemple Dreesbeke *et al.* [1989], Gouriéroux et Monfort [1990] et Mélard [1990]. Dans les chapitres suivants, des processus à temps continu seront aussi considérés.

Nous donnons ici les principales définitions relatives aux processus stochastiques (Gouriéroux [1992]), en particulier les processus stationnaires et les processus autorégressifs-moyenne mobile (« autoregressive moving average » ou ARMA).

2.2.1 Processus stochastiques

Un processus est entièrement défini par l'ensemble des distributions finies des variables aléatoires, $(Y_{t_1}, \dots, Y_{t_k})$, pour tout k et pour tout k -uplet (t_1, \dots, t_k) . En supposant des lois continues, la densité $f_{t_1, \dots, t_k}(y_{t_1}, \dots, y_{t_k})$ peut être évaluée à l'aide des densités marginales et conditionnelles, par exemple la densité marginale de Y_{t_1} , la distribution conditionnelle de $Y_{t_2} | Y_{t_1}$ (distribution de Y_{t_2} conditionnellement à Y_{t_1}), et ainsi de suite :

$$f_{t_1, \dots, t_k}(y_{t_1}, \dots, y_{t_k}) = f_{t_1}(y_{t_1}) f_{t_2|t_1}(y_{t_2} | y_{t_1}) \dots f_{t_k|t_1, \dots, t_{k-1}}(y_{t_k} | y_{t_1}, \dots, y_{t_{k-1}}). \quad (2.1)$$

Définition 2.1

Un processus Y est dit *du second ordre* si et seulement si les deux conditions suivantes sont remplies :

$$\begin{aligned} E(Y_t) &= m_t \text{ existe } \forall t; \\ \text{Cov}(Y_t, Y_s) &= E[(Y_t - m_t)(Y_s - m_s)'] \text{ existe } \forall t, s. \end{aligned}$$

Considérons le cas particulier d'un *processus gaussien*. Cela signifie que la distribution de $(Y_{t_1}, \dots, Y_{t_k})$ est normale, pour tout k et pour tout k -uplet

(t_1, \dots, t_k) . Toutes les distributions marginales et conditionnelles d'un processus gaussien sont aussi normales. Rappelons que la distribution conditionnelle d'une distribution normale a comme moyenne une fonction affine des valeurs. Par exemple $f_{t_k|t_1, \dots, t_{k-1}}(y_{t_1}, \dots, y_{t_{k-1}})$ a comme moyenne l'espérance conditionnelle, $E(Y_{t_k}|Y_{t_1}, \dots, Y_{t_{k-1}})$, qui est aussi la régression linéaire (théorique) de Y_{t_k} en fonction de $Y_{t_1}, \dots, Y_{t_{k-1}}$, $EL(Y_{t_k}|Y_{t_1}, \dots, Y_{t_{k-1}})$, évaluée en $Y_{t_1} = y_{t_1}, \dots, Y_{t_{k-1}} = y_{t_{k-1}}$. La variance conditionnelle est constante par rapport à $(y_{t_1}, \dots, y_{t_{k-1}})$ et vaut la variance résiduelle associée à la même régression. Un processus gaussien est un processus du second ordre.

Dans le cas d'un processus du second ordre non gaussien, l'espérance conditionnelle $E(Y_{t_2}|Y_{t_1})$ n'est pas nécessairement linéaire en Y_{t_1} . C'est la meilleure approximation de Y_{t_2} , au sens de la moyenne quadratique, comme fonction arbitraire de Y_{t_1} . On introduit également l'espérance conditionnelle linéaire $EL(Y_{t_2}|Y_{t_1})$ qui correspond à la meilleure approximation de Y_{t_2} , au sens de la moyenne quadratique, comme fonction linéaire affine de Y_{t_1} . Cette notion correspond donc à la régression linéaire. Dans le cas gaussien, espérance conditionnelle et espérance conditionnelle linéaire sont confondues.

Pour un processus multidimensionnel Y , on peut considérer les composantes Y_{jt} , $j = 1, \dots, n$, de Y_t . Chaque série $(Y_{jt}, t \in \mathbb{Z})$ constitue un processus unidimensionnel Y_j , $j = 1, \dots, n$.

2.2.2 Processus stationnaires

Dans le contexte des modèles linéaires, on considère des processus dits stationnaires.

Définition 2.2

Un processus Y est faiblement stationnaire (ou stationnaire du second ordre) si et seulement si les deux conditions suivantes sont remplies :

$$E(Y_t) = m, \forall t;$$

$$\text{Cov}(Y_t, Y_{t+h}) = \Gamma(h), \forall t, \forall h.$$

Un processus faiblement stationnaire est de moyenne constante et homoscédastique. $\Gamma(\cdot)$ est appelé la fonction d'autocovariance. En divisant chaque autocovariance par le produit des écart-types, on obtient les autocorrélations qui constituent la fonction d'autocorrélation $P(\cdot)$, ou $\rho(\cdot)$ dans le cas unidimensionnel.

Définition 2.3

Un processus Y est fortement stationnaire si et seulement si la distribution de $(Y_{t_1+h}, Y_{t_2+h}, \dots, Y_{t_k+h})$ est égale à la distribution de $(Y_{t_1}, Y_{t_2}, \dots, Y_{t_k})$:

$$f_{t_1+h, \dots, t_k+h}(y_{t_1+h}, \dots, y_{t_k+h}) = f_{t_1, \dots, t_k}(y_{t_1}, \dots, y_{t_k}),$$

pour tout h , pour tout k et pour tout k -uple (t_1, \dots, t_k) .

Un processus du second ordre qui est fortement stationnaire est également faiblement stationnaire. Un processus gaussien faiblement stationnaire est aussi fortement stationnaire.

Les séries traitées peuvent rarement être considérées comme générées par des processus stationnaires. Des transformations permettent de rendre constantes la moyenne et la variance : par exemple, la différenciation (ΔY_t ou $\nabla Y_t = Y_t - Y_{t-1}$) pour la moyenne, et la transformation logarithmique pour la variance.

L'étude de la dépendance entre les variables d'un processus Y , compte tenu de la structure temporelle, peut être effectuée par l'espérance conditionnelle $E(Y_t | Y_{t-1}, Y_{t-2}, \dots)$ ou par l'espérance conditionnelle linéaire $EL(Y_t | Y_{t-1}, Y_{t-2}, \dots)$. Au temps t notons le passé du processus par $\underline{Y}_{t-1} = (Y_{t-1}, Y_{t-2}, \dots)$. Rappelons une propriété importante de l'espérance conditionnelle, qu'on appelle théorème des projections itérées (et qui généralise le théorème des trois perpendiculaires) :

$$E(Y_{t+h} | \underline{Y}_{t-1}) = E(E(Y_{t+h} | Y_t) | \underline{Y}_{t-1}), \quad (2.2)$$

où h est un entier strictement positif. Ce théorème s'applique également à l'espérance conditionnelle linéaire.

2.2.3 Processus bruit blanc

Le processus le plus simple est le *bruit blanc* (« white noise »). Comme ci-dessus, deux définitions sont naturelles, l'une forte, l'autre faible, pour ce processus, noté $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$.

Définition 2.4

ε est un processus *bruit blanc fort* si et seulement si les ε_t sont des variables aléatoires indépendantes et identiquement distribuées (i.i.d.).

Définition 2.5

ε est un processus *bruit blanc faible* si et seulement si les deux conditions suivantes sont remplies :

$$\begin{aligned} EL(\varepsilon_t | \underline{\varepsilon}_{t-1}) &= 0, & \forall t; \\ V(\varepsilon_t) &= \Omega, & \forall t. \end{aligned}$$

Un processus bruit blanc faible qui est gaussien est aussi un processus bruit blanc fort. D'autres définitions peuvent aussi être introduites.

Définition 2.6

ε est une *différence de martingale homoscédastique* si et seulement si les deux conditions suivantes sont remplies :

$$\begin{aligned} E(\varepsilon_t | \underline{\varepsilon}_{t-1}) &= 0; & \forall t; \\ V(\varepsilon_t) &= \Omega, & \forall t. \end{aligned}$$

Si l'on supprime l'exigence que la variance soit constante dans chacune de ces définitions, on ajoute le qualificatif *hétéroscédastique* (au lieu d'homoscédastique). En particulier, nous utiliserons fréquemment la définition suivante.

Définition 2.7

ε est une *différence de martingale conditionnellement hétéroscédastique* si et seulement si les deux conditions suivantes sont remplies :

$$\begin{aligned} E(\varepsilon_t | \underline{\varepsilon}_{t-1}) &= 0; & \forall t; \\ V(\varepsilon_t | \underline{\varepsilon}_{t-1}) &= \Omega_t, & \forall t. \end{aligned}$$

2.2.4 Classes de processus

Définition 2.8

Un processus *régulier* (on dit aussi purement indéterminable) est un processus tel que $\lim_{h \rightarrow \infty} E(Y_t | Y_{t-h}) = E(Y_t)$, pour tout t , c'est-à-dire que le passé lointain est sans influence sur le comportement présent.

Dans la suite, on désigne par L l'*opérateur de retard* (noté aussi B ou q^{-1}), tel que $LY_t = Y_{t-1}$. On peut de même introduire comme opérateur un polynôme en L , par exemple l'*opérateur de différence* $\nabla = 1 - L$. On peut aussi introduire comme opérateur une série en L , formellement tout au moins. Par exemple

$$\frac{1}{1 - L} = 1 + L + L^2 + \dots$$

est l'*opérateur d'accumulation*, inverse de l'*opérateur de différence*.

Théorème 2.1 (décomposition de Wold)

Un processus Y faiblement stationnaire régulier admet une représentation

$$Y_t = \varepsilon_t + A_1 \varepsilon_{t-1} + \dots + A_j \varepsilon_{t-j} + \dots = A(L)\varepsilon_t, \tag{2.3}$$

où les $\varepsilon_t = Y_t - E(Y_t | \underline{\varepsilon}_{t-1})$ constituent le **processus innovation** qui est un *processus bruit blanc faible*.

Les coefficients A_j satisfont la condition

$$\sum_{j=1}^{\infty} A_j \Omega A_j' < \infty$$

qui exprime que la série aléatoire (2.3) converge au sens de la moyenne quadratique.

On peut donner une idée de la démonstration de ce théorème. Considérons l'espace des variables aléatoires du second ordre généré par $(Y_s, s \leq t)$, soit E_t . Munissons-le de la covariance comme produit scalaire. La norme d'une variable aléatoire est alors l'écart-type et l'orthogonalité correspond à la non-corrélation (l'espace est alors appelé espace de Hilbert). Projétons orthogonalement Y_t sur E_{t-1} , le sous-espace généré par $(Y_s, s \leq t-1)$, et notons ε_t la différence entre Y_t et la projection qui n'est autre que $EL(Y_t, |Y_{t-1})$. Donc $Y_t = \varepsilon_t + EL(Y_t, |Y_{t-1})$. Cette projection appartenant à E_{t-1} , peut s'écrire sous la forme $A_1 Y_{t-1}$ plus un élément de E_{t-2} , et peut donc être décomposée, de même, en $A_1 \varepsilon_{t-1}$ plus un élément de E_{t-2} . Par conséquent, $Y_t = \varepsilon_t + A_1 \varepsilon_{t-1}$ plus un élément de E_{t-2} , et ainsi de suite. La restriction à un processus régulier revient à exprimer que l'espace E_t se réduit à $\{0\}$ quand $t \rightarrow -\infty$.

Remarquons que la décomposition de Wold peut se généraliser à un processus du second ordre régulier, même s'il n'est pas faiblement stationnaire (par exemple, Mélard, [1985], chapitre 2). On parle alors de décomposition de Wold-Cramér. Les différences par rapport à (2.3) sont que les coefficients dépendent alors du temps, doivent donc s'écrire A_{tj} , et que le processus ε est un bruit blanc faible hétéroscédastique, tel que $V(\varepsilon_t) = \Omega_t$.

Les modèles que nous considérons plus loin sont définis par une équation aux différences stochastique de la forme

$$Y_t = \varepsilon_t + g(Y_{t-1}, Y_{t-2}, \dots, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots), \quad (2.4)$$

où ε est un processus bruit blanc donné. La première question est l'existence et l'unicité d'une solution stationnaire de l'équation du modèle. On demande à ce processus d'être causal c'est-à-dire que les Y_t peuvent s'exprimer en fonction des ε_s , présent et passés.

Définition 2.9

Y est un processus *causal* si et seulement s'il existe une fonction mesurable h telle que $Y_t = h(\varepsilon_t, \varepsilon_{t-1}, \dots)$, pour tout t .

Pour certains besoins (prévision, évaluation de la fonction de vraisemblance, certains tests), on souhaite résoudre le problème inverse, c'est-à-dire retrouver les ε_t , les Y_t vérifiant l'équation (2.4) étant donnés. C'est ce qu'on appelle l'*inversibilité* du processus.

Définition 2.10

Y est un processus *inversible* si et seulement s'il existe une fonction mesurable h telle que $\varepsilon_t = h(Y_t, Y_{t-1}, \dots)$, pour tout t .

Par exemple le processus unidimensionnel Y défini par

$$Y_t = \varepsilon_t + 2\varepsilon_{t-1}, \tag{2.5}$$

où ε est un bruit blanc faible, n'est pas inversible car on aurait une série non convergente

$$\varepsilon_t = Y_t + 2Y_{t-1} + 4Y_{t-2} + \dots$$

(on peut d'ailleurs montrer que (2.5) n'est la représentation de Wold d'aucun processus).

Un cas particulier de (2.3) est celui où les coefficients A_j de la décomposition de Wold sont nuls à partir d'une valeur finie q de j . On parle alors de processus $MA(q)$ ou $ARMA(0, q)$.

Définition 2.11

Un processus *moyenne mobile d'ordre q* ou $MA(q)$ est défini par l'équation suivante :

$$Y_t = (I - \theta_1 L - \dots - \theta_q L^q) \varepsilon_t. \tag{2.6}$$

La condition pour qu'un processus moyenne mobile soit inversible est que les racines du polynôme $\det(\theta(z))$ soient toutes de module supérieur à 1, où

$$\theta(L) = I - \theta_1 L - \dots - \theta_q L^q$$

est le *polynôme moyenne mobile* de degré q . Une caractéristique des processus $MA(q)$ est qu'ils sont q -dépendants, c'est-à-dire que leur fonction d'auto-corrélation $P(h)$ est nulle pour $h > q$.

A part la décomposition de Wold, un autre moyen d'introduire une sous-classe de processus stochastiques est de se limiter aux processus markoviens.

Définition 2.12

Un processus est *fortement markovien d'ordre p* si et seulement si

$$f_{t|t-1, \dots, t-h}(y_t | y_{t-1}, \dots, y_{t-h}) = f_{t|t-1, \dots, t-p}(y_t | y_{t-1}, \dots, y_{t-p}), \quad \forall h \geq p,$$

pour tout t .

Toute l'information d'un processus fortement markovien d'ordre p est contenue dans les p valeurs les plus récentes. Par analogie avec les notions introduites pour le bruit blanc, on peut définir un processus conditionnellement markovien ou faiblement markovien, en recourant à l'espérance conditionnelle ou à l'espérance conditionnelle linéaire.

Définition 2.13

Un processus est *conditionnellement markovien d'ordre p* si et seulement si

$$E(Y_t | Y_{t-1}, Y_{t-2}, \dots) = E(Y_t | Y_{t-1}, \dots, Y_{t-p}),$$

pour tout t .

Définition 2.14

Un processus est *conditionnellement linéairement markovien d'ordre p* si et seulement si

$$EL(Y_t | Y_{t-1}, Y_{t-2}, \dots) = EL(Y_t | Y_{t-1}, \dots, Y_{t-p})$$

pour tout t .

Un processus conditionnellement linéairement markovien d'ordre p est aussi appelé *autorégressif d'ordre p* , ou $AR(p)$. Il vérifie une équation du type

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

ou

$$(1 - \phi_1 L - \dots - \phi_p L^p) Y_t = \varepsilon_t. \quad (2.7)$$

Pour avoir la stationnarité, une condition nécessaire sur les coefficients ϕ_j est que les racines du polynôme $\det(\phi(z))$ soient toutes de module strictement supérieur à 1, où

$$\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$$

est le *polynôme autorégressif* de degré p .

Plus généralement, nous parlerons de *processus mixte autorégressif-moyenne mobile d'ordre p et q* , ou $ARMA$ d'ordre (p, q) , ou $ARMA(p, q)$ si Y vérifie l'équation

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

ou

$$(1 - \phi_1 L - \dots - \phi_p L^p) Y_t = (1 - \theta_1 L - \dots - \theta_q L^q) \varepsilon_t. \quad (2.8)$$

Les innovations ε constituent un processus bruit blanc, dans un des sens spécifiés ci-dessus.

2.2.5 Etude des processus ARMA

Les processus $ARMA$ servent souvent de modèles pour décrire l'évolution de séries temporelles. On parle alors de *modèles ARMA*. Cela signifie qu'un

processus ARMA déterminé sera considéré comme ayant généré la série temporelle donnée ou encore que celle-ci pourra être vue comme une réalisation de ce processus ARMA.

Nous considérons quelques exemples de processus ARMA, illustrés par des séries temporelles construites à partir d'une série générée par un processus bruit blanc gaussien, dont les variables ε_t sont de moyenne nulle et de variance 1. La figure 2.1 montre la série et les autocorrélations de la série des 100 premières valeurs. Les intervalles de signification asymptotiques à 5% sur les autocorrélations, déterminés par la formule $\pm 1.96(T)^{1/2}$, sont présentés dans les graphiques des autocorrélations.

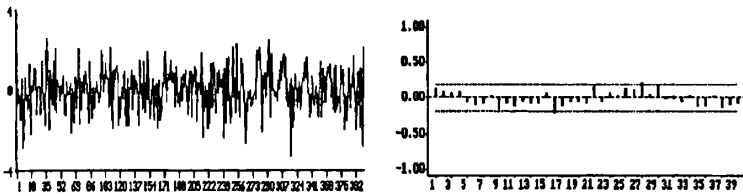


Figure 2.1

Graphique de la série de longueur 400 générée par un processus bruit blanc et les autocorrélations de la série des 100 premières données

- 1) Le processus AR(1), d'équation $Y_t = 5 + 0.8(Y_{t-1} - 5) + \varepsilon_t$, (voir figure 2.2).

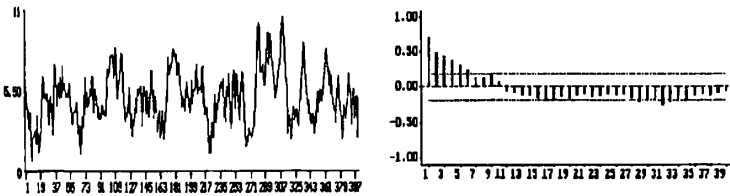


Figure 2.2

Graphique de la série de longueur 400 générée par un processus AR(1) et les autocorrélations de la série des 100 premières données

- 2) Le processus MA(1), d'équation $Y_t = \varepsilon_t - 0.5\varepsilon_{t-1}$ (voir figure 2.3).
- 3) Le processus ARMA(1,1), d'équation $Y_t - 0.9Y_{t-1} = \varepsilon_t - 0.5\varepsilon_{t-1}$ (voir figure 2.4).

Il est possible, dans certains cas, de rendre stationnaire une série non stationnaire. Nous avons déjà constaté que la différence première d'une série temporelle non stationnaire peut être stationnaire. Nous montrons ici comment générer une série non stationnaire à partir d'une série stationnaire.

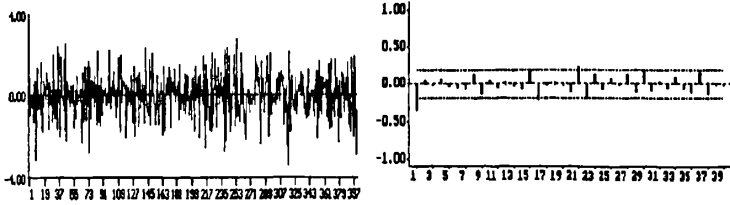


Figure 2.3

Graphique de la série de longueur 400 générée par un processus MA(1) et les autocorrélations de la série des 100 premières données

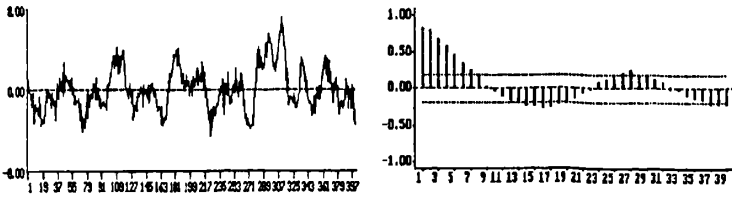


Figure 2.4

Graphique de la série de longueur 400 générée par un processus ARMA(1,1) et les autocorrélations de la série des 100 premières données

A partir de la série artificielle de la figure 2.4 on obtient une série avec variations de niveau en définissant $y_1 = Y_1$ et $y_t = y_{t-1} + Y_t$, pour $t > 1$. Le modèle est défini par l'équation

$$(1 - 0.9L)\nabla y_t = (1 - 0.5L)\varepsilon_t.$$

La série obtenue est illustrée dans la figure 2.5 (extraite de Mélard [1990]).

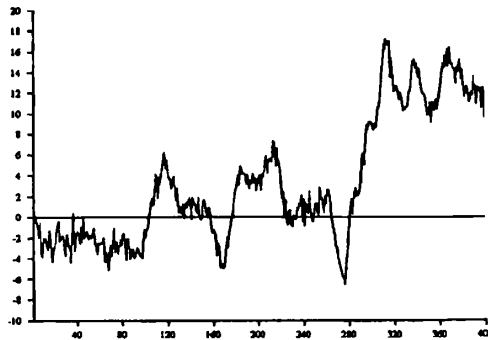


Figure 2.5: Série artificielle avec variations de niveau

Une autre possibilité est d'employer une transformation instantanée de la

variable. Elle permet de rendre stationnaire une série de dispersion variable, liée au niveau atteint par la variable.

La figure 2.6 (extraite de Mélard [1990]) présente une série artificielle générée comme suit. Notons $\nabla_4 = 1 - L^4$. Nous nous sommes basé sur une série artificielle générée par le modèle suivant (parfois appelé SARIMA (0, 1, 1) (0, 1, 1)₄)

$$\nabla_4 Y_t = (1 - 0.6L)(1 - 0.8L^4)\varepsilon_t,$$

divisée par 50, afin d'en réduire l'ordre de grandeur. Ensuite, nous avons posé $y_t = e^{Y_t}$.

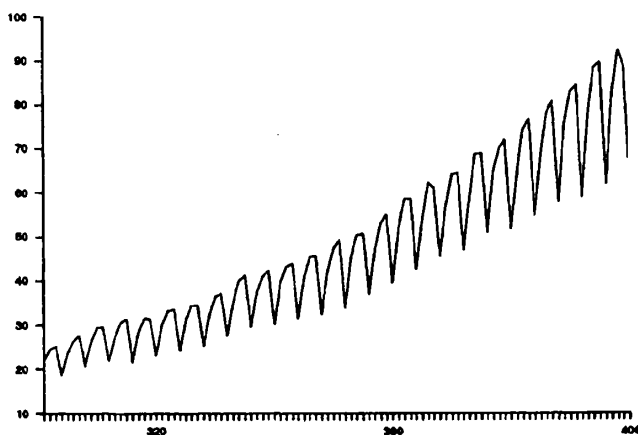


Figure 2.6: Série artificielle qui demande à être transformée en logarithmes

2.3 MODÈLES LINÉAIRES

2.3.1 Description de la méthode de Box et Jenkins

Nous décrivons dans ce paragraphe, de manière succincte, l'élaboration d'un modèle linéaire de type ARMA. On distingue trois étapes : la *spécification* du modèle (ou *identification*) ; l'*ajustement* du modèle (ou *estimation*) ; la *validation* du modèle (ou *adéquation*). Ces trois étapes sont généralement répétées jusqu'à l'obtention d'un modèle satisfaisant qui peut alors être employé.

Il faut insister sur le besoin de se familiariser avec les données, et plus généralement les théories qui existent dans le domaine dont elles relèvent. Il est important de connaître les événements qui ont éventuellement influencé la série et donné naissance à des données aberrantes du point de vue statistique. On peut à ce stade décider de ne traiter qu'une partie de la série, de corriger des données, d'effectuer une transformation de variable.

La classe de modèles est basée sur le modèle ARMA mais comporte aussi les opérateurs de différence, *ordinaire* ($\nabla = 1 - L$) ou *saisonniers*

($\nabla_s = 1 - L^s$), une constante et la possibilité de *transformation de puissance* de la variable. On utilise plus précisément $(Y_t^\lambda - 1)/\lambda$ pour retrouver la transformation logarithmique comme cas particulier, lorsque λ tend vers 0. L'équation du modèle général est la suivante :

$$\phi(L)\Phi(L^s) \left[\nabla^d \nabla_s^D \left(\frac{Y_t^\lambda - 1}{\lambda} \right) \right] = C + \theta(L)\Theta(L^s)\varepsilon_t.$$

Notons comme précédemment p et q les degrés de $\phi(L)$ et $\theta(L)$, les polynômes ordinaires. Notons p' et q' , les degrés des polynômes en L^s , respectivement $\Phi(L^s)$ et $\Theta(L^s)$, les polynômes saisonniers. Les différences doivent rendre la série stationnaire en niveau. Le choix est souvent effectué sur base de graphiques. Le choix de la transformation des données est plus facile quand on applique les opérateurs de différence adéquats.

La spécification du modèle se base sur les fonctions d'autocorrélation et d'autocorrélation partielle de la série différenciée. L'*autocorrélation de retard* h , $\rho(h)$, est estimée par

$$\hat{\rho}(h) = \frac{\sum_{t=h+1}^T (Y_t - \bar{Y})(Y_{t-h} - \bar{Y})}{\sum_{t=1}^T (Y_t - \bar{Y})^2}, \quad (2.9)$$

où \bar{Y} est la moyenne de la série de 1 à T . On tient compte du fait que, sous l'hypothèse d'un bruit blanc gaussien, la distribution asymptotique de $T^{1/2}\hat{\rho}(h)$ est $N(0, 1)$.

L'*autocorrélation partielle de retard* h , $\pi(h)$, d'un processus stationnaire est le coefficient de corrélation partielle entre Y_t et Y_{t-h} en éliminant l'influence de $Y_{t-1}, \dots, Y_{t-h+1}$. La suite des $\pi(h)$ définit la *fonction d'autocorrélation partielle*. La particularité de celle-ci est d'être tronquée pour un processus autorégressif. Plus précisément, pour un processus AR(p), $\pi(h) = 0$ pour $h > p$. On peut estimer les $\pi(h)$ par l'autorégression suivante :

$$Y_t = \phi_1^{(h)} Y_{t-1} + \dots + \phi_h^{(h)} Y_{t-h} + \varepsilon_t,$$

où $\hat{\pi}(h)$ est l'estimation de $\phi_h^{(h)}$. Sous l'hypothèse d'un bruit blanc gaussien, la loi asymptotique de $T^{1/2}\hat{\pi}(h)$ est $N(0, 1)$. Aussi bien pour les autocorrélations que pour les autocorrélations partielles, les limites de signification à 5% sont $\pm 1.96(T)^{-1/2}$.

La spécification d'un modèle autre que AR ou MA est plus délicate. On propose souvent (par exemple Mélard, [1990]) de procéder par étape et d'introduire un terme à la fois dans les polynômes MA et AR, ordinaires et saisonniers, correspondant à une autocorrélation significative ou une autocorrélation partielle significative, pour un retard que l'on peut expliquer (1, 2,

peut-être 3, 4 dans le cas de données trimestrielles, 12 dans le cas de données mensuelles, etc.).

Il existe d'autres méthodes de spécification de modèles (fonction d'auto-corrélation inverse, « extended autocorrelation function », méthode du coin, etc., De Gooijer *et al.*, [1985]) qui proposent généralement plusieurs solutions. Une approche simple et efficace, appelée la *procédure de spécification autorégressive* par Mélard [1990] et basée sur les travaux de Tiao et Box [1981] et de Tsay et Tiao [1984], consiste à estimer successivement les paramètres de modèles AR(1), AR(2), et ainsi de suite sur les données $\{Y_t\}$, et à examiner les autocorrélations des séries résiduelles. Si l'ajustement d'un modèle AR(p) conduit à une série résiduelle dont les autocorrélations sont tronquées au-delà du retard q , cette série résiduelle pourra être représentée par un modèle MA(q), d'où ressort un modèle global ARMA(p, q) pour la série $\{Y_t\}$.

L'estimation des paramètres est réalisée par la *méthode du maximum de vraisemblance*, sous l'hypothèse d'un processus gaussien stationnaire et inversible. Pour la simplicité, considérons le modèle de la forme

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}.$$

Box et Jenkins [1976] ont proposé une *méthode des moindres carrés conditionnelle*, où pour amorcer les relations de récurrence,

$$\varepsilon_t = Y_t - \phi_1 Y_{t-1} - \dots - \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q},$$

on posait égales à zéro les q premières erreurs, $\varepsilon_p, \dots, \varepsilon_{p+1-q}$ de manière à calculer

$$\varepsilon_{p+1} = Y_{p+1} - \phi_1 Y_p - \dots - \phi_p Y_1.$$

La somme des carrés à minimiser

$$\sum_{t=p+1}^T \varepsilon_t^2$$

n'étant pas une fonction quadratique des paramètres $\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$, il faut recourir à une procédure numérique itérative de *régression non linéaire ou d'optimisation d'une fonction*. La convergence n'est pas assurée, peut dépendre des valeurs initiales et peut survenir en un extremum local. Il n'y a généralement pas de problème quand le modèle est bien choisi. Box et Jenkins ont proposé une méthode de prévision à rebours (« backforecasting ») qui améliore les conditions initiales.

De plus en plus, on emploie la fonction de vraisemblance exacte pour laquelle il existe maintenant des algorithmes performants (par exemple, Mélard [1984]). La fonction de vraisemblance est la densité des observations, considérée comme fonction des paramètres. Pour évaluer cette densité, on emploie (2.1) et la décomposition de Wold-Cramér du processus $(Y_t, t \geq 1)$, en effectuant

un changement de variable de Y_t vers l'innovation « estimée » $\tilde{\varepsilon}_t$, normale de moyenne 0 et de variance $\sigma^2 h_t^2$, où $\tilde{\varepsilon}_t$ et h_t dépendent de β . Dès lors, la fonction de vraisemblance s'écrit

$$L(Y; \beta) = (2\pi\sigma)^{-T/2} \left(\prod_{t=1}^T h_t^2 \right)^{-1/2} \exp \left[-\frac{1}{2\sigma^2} \sum_{t=1}^T \frac{\tilde{\varepsilon}_t^2}{h_t^2} \right].$$

Le calcul des innovations estimées peut se faire par le filtre de Kalman rapide, étant donné la stationnarité du processus. En concentrant la vraisemblance par rapport à σ , le problème peut se ramener à minimiser numériquement une somme de carrés

$$\sum_{t=1}^T \left\{ \frac{\tilde{\varepsilon}_t}{h_t} \left(\prod_{s=1}^T h_s \right)^{1/T} \right\}^2,$$

ce qui fournit l'estimateur $\hat{\beta}$ et les résidus normalisés correspondants $\hat{\varepsilon}_t/\hat{h}_t$. Ensuite, l'estimation de σ^2 s'obtient par

$$\hat{\sigma}^2 = \sum_{t=1}^T \left(\frac{\hat{\varepsilon}_t}{\hat{h}_t} \right)^2.$$

Asymptotiquement, ces différentes méthodes sont équivalentes. De plus, les estimateurs sont convergents et asymptotiquement normaux, sans que l'hypothèse d'un processus gaussien soit nécessaire, moyennant des suppositions de régularité assez générales sur le processus (Gourieroux *et al.* [1984]).

Les procédures numériques fournissent une estimation de la matrice de covariance asymptotique des estimateurs, ce qui permet de réaliser des tests de signification des coefficients du modèle en considérant un rapport de Student.

L'adéquation du modèle (ou validation) consiste essentiellement en une *analyse des résidus* $\hat{\varepsilon}_t$. On emploie surtout les autocorrélations résiduelles

$$\hat{\rho}_{\hat{\varepsilon}}(h) = \frac{\sum_{t=h+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-h}}{\sum_{t=1}^T \hat{\varepsilon}_t^2}.$$

On teste que le modèle est correct au moyen de la *statistique Q de Box et Pierce*, de la forme

$$Q = T \sum_{h=1}^H \hat{\rho}_{\hat{\varepsilon}}^2(h), \quad (2.10)$$

où H est le nombre d'autocorrélations utilisées (ni trop grand par rapport à T , ni trop petit). Sous l'hypothèse que le modèle est correct, la distribution

asymptotique de Q est χ^2_{H-p-q} , khi carré à $H - p - q$ degrés de liberté (c'est-à-dire le nombre d'autocorrélations utilisées diminué du nombre de paramètres estimés). La statistique Q' de Ljung et Box, définie par

$$Q' = T(T + 2) \sum_{h=1}^H \frac{\hat{\rho}_{\hat{\epsilon}}^2(h)}{T - h} \tag{2.11}$$

est préférable pour des séries courtes.

Si le modèle ne comporte pas de constante, il y a lieu de tester l'hypothèse que la moyenne des erreurs est nulle. Il est aussi recommandé d'examiner la présence de *données aberrantes*. Deux données aberrantes distantes de h induisent de l'autocorrélation de retard h . Il faut éventuellement corriger les données ou utiliser *l'analyse d'interventions*. Si le modèle est rejeté pour l'une ou l'autre raison, il faut reprendre l'analyse à l'étape de spécification en incorporant l'information qui a été acquise. Le modèle complété peut alors être estimé. L'analyse des résidus de ce modèle est effectuée, et ainsi de suite jusqu'à l'obtention d'un modèle satisfaisant. Si plusieurs modèles satisfaisants sont trouvés, ils peuvent être comparés entre eux à l'aide d'un *critère d'information* comme

$$AIC = \log(\hat{\sigma}^2) + \frac{2m}{T}$$

ou mieux

$$SBIC = \log(\hat{\sigma}^2) + \frac{m \log(T)}{T},$$

où m est le nombre de paramètres. Notons que SBIC pénalise plus les paramètres inutiles. Ces critères doivent être minimisés.

La modélisation sera illustrée sur deux exemples, réalisés à l'aide de MicroTSP 5.1. Dans les tableaux donnant les autocorrélations, celles-ci sont représentées dans le graphique de gauche et les valeurs en sont imprimées dans la colonne « ac », en fonction du retard indiqué dans la colonne précédente ; de même pour les autocorrélations partielles, dans le graphique de droite et la colonne « pac ».

2.3.2 Exemple 1

Le premier exemple est relatif au produit intérieur brut de l'Italie aux prix du marché. La variable est notée PIB. Les données couvrent la période allant du 1er trimestre de 1980 au 4e trimestre de 1991. L'étude porte sur la période allant du 1er trimestre de 1980 au 4e trimestre de 1989 afin de laisser deux années pour la comparaison des données aux prévisions. Les données sont présentées dans la figure 2.7.

La série n'est pas stationnaire puisqu'elle présente une tendance linéaire marquée. On a donc considéré DPIB, la série du PIB en différences premières ordinaires pour laquelle le tableau 2.1 présente les autocorrélations et les autocorrélations partielles.

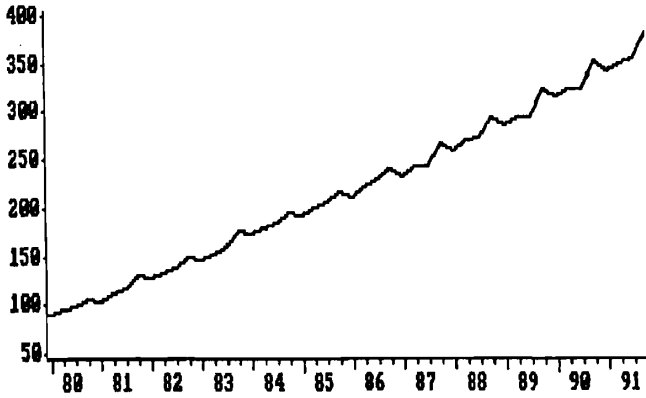


Figure 2.7: Les données du produit intérieur brut

Tableau 2.1: Autocorrélations et autocorrélations partielles de la série DPIP

Autocorrélations		Partial Autocorrélations		ac	pac	
*****	*****	*****	.	1	-0.605	-0.605
*****	*****	*****	.	2	0.380	0.022
*****	*****	*****	*****	3	-0.640	-0.635
*****	*****	.	*****	4	0.769	0.394
*****	*****	.	**	5	-0.465	-0.132
*****	*****	.	**	6	0.333	-0.157
*****	*****	.	**	7	-0.535	0.065
*****	*****	.	**	8	0.607	0.323
*****	*****	.	.	9	-0.334	0.097
*****	*****	.	.	10	0.258	-0.021
*****	*****	.	**	11	-0.404	-0.156
*****	*****	.	*	12	0.399	-0.063
Q-Statistic (12 lags) 116.778		S.E. of Correlations 0.160				

La présence périodique de fortes autocorrélations suggère d'appliquer une différence saisonnière. Notons $XPIB = \nabla \nabla_4 PIB$. La figure 2.8 montre que la série a été approximativement rendue stationnaire.

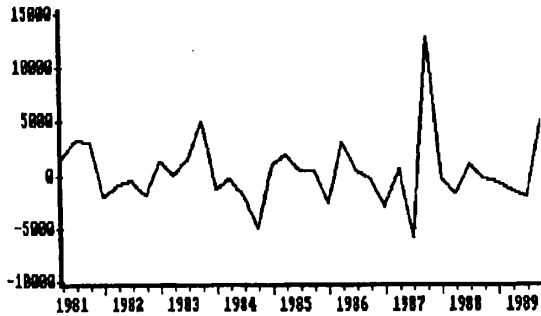


Figure 2.8: Les données de $\nabla \nabla_4 PIB$

Tableau 2.2: Autocorrélations et autocorrélations partielles de la série XPIB

Autocorrélations		Partial Autocorrélations		ac	pac	
***	.	***	.	1	-0.228	-0.228
. **	.	. **	.	2	-0.033	-0.090
. **	.	. **	.	3	-0.116	-0.155
.	.	. **	.	4	-0.180	-0.273
.	.	. **	.	5	-0.013	-0.185
***	*	. **	.	6	-0.053	-0.088
***	****	. **	.	7	-0.208	-0.380
.	****	. **	.	8	0.317	0.047
.	****	.	.	9	-0.003	-0.009
.	**	.	*	10	0.129	0.115
.	*	.	***	11	0.093	0.234
***	.	.	.	12	-0.257	-0.019
O-Statistic (12 lags)		11.803		S.E. of Correlations		0.169

Sur la base du tableau 2.2, on ne constate aucune autocorrélation significative au niveau de 5%. La statistique Q de Box et Pierce (à comparer à la loi χ^2 à 12 degrés de liberté) conduit à l'acceptation d'un modèle très simple :

$$\nabla\nabla_4\text{PIB}_t = C + \varepsilon_t$$

ou

$$\text{PIB}_t = C + \text{PIB}_{t-1} + \text{PIB}_{t-4} - \text{PIB}_{t-5} + \varepsilon_t.$$

Ce modèle a été estimé dans le tableau 2.3 en employant la méthode des moindres carrés. La constante n'est même pas significative. Les prévisions sont

$$\widehat{\text{PIB}}_T(h) = \widehat{\text{PIB}}_T(h - 1) + \widehat{\text{PIB}}_T(h - 4) - \widehat{\text{PIB}}_T(h - 5),$$

où $\widehat{\text{PIB}}_T(h) = \text{PIB}_{t+h}$, quand $t + h \leq T$.

Tableau 2.3: Estimation d'un modèle simple sur la série XPIB

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	354.08571	548.42919	0.6456362	0.523
R-squared	0.000000	Mean of dependent var	354.0857	
Adjusted R-squared	0.000000	S.D. of dependent var	3244.551	
S.E. of regression	3244.551	Sum of squared resid	3.58D+08	
Durbin-watson stat	2.390963	Log likelihood	-332.1212	

On peut constater dans la figure 2.9 que les prévisions pour 1990 et 1991 sont excellentes, presque parfaites. Etant donné qu'un écart-type des innovations égal à 3 245 a été estimé dans le tableau 2.3, l'intervalle de prévision à 95% pour le premier trimestre de 1990 s'obtient en ajoutant et retranchant $(1.96)(3\ 245)$ à la prévision. Pour les horizons suivants, on se base sur l'écriture du modèle sous forme de *moyenne mobile infinie*

$$\begin{aligned} \text{PIB}_t &= (\nabla\nabla_4)^{-1} \varepsilon_t \\ &= (1 + L + L^2 + \dots)(1 + L^4 + L^8 + \dots)\varepsilon_t \\ &= (1 + L + L^2 + L^3 + 2L^4 + \dots)\varepsilon_t. \end{aligned}$$

Les variances pour les horizons 2, 3, 5, etc., sont respectivement multipliées par $1 + 1 = 2$, $1 + 1 + 1 = 3$, $1 + 1 + 1 + 1 + 2^2 = 8$, etc. Plus généralement, le

facteur par lequel la variance est multipliée pour l'horizon h est la somme des carrés des h premiers termes de la forme MA (∞) du processus.

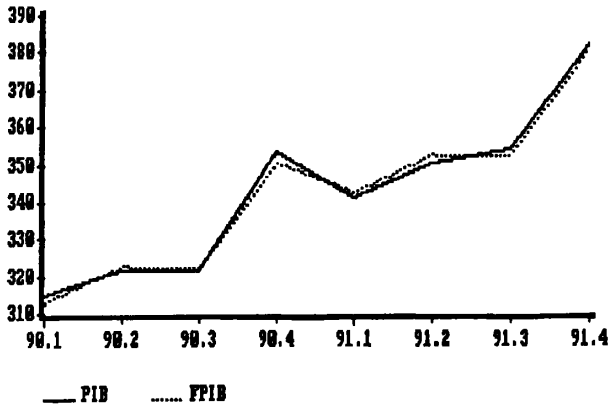


Figure 2.9: Les données de PIB et de ses prévisions FPIB

2.3.3 Exemple 2

Le deuxième exemple concerne les prix de la viande de taureau («cul de polain»), de janvier 1975 à décembre 1990. Les données de 1991 et du début de 1992 ne sont pas utilisées pour le choix du modèle. On a considéré les séries suivantes :

- TAUR : la série des prix ;
- DTAUR : la série TAUR en différences premières ordinaires.

Sur base de la figure 2.10, on comprend qu'il sera plus facile de modéliser la série DTAUR.

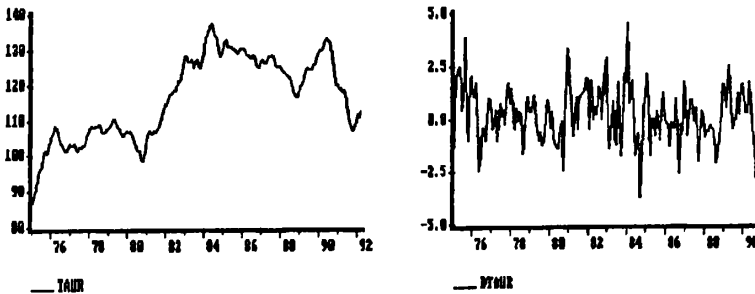


Figure 2.10: Les séries TAUR et DTAUR

Le tableau 2.4 montre qu'on peut envisager deux possibilités pour commencer. On considère soit que les autocorrélations partielles sont tronquées après

Tableau 2.4: Autocorrélations et autocorrélations partielles de la série DTAUR

Autocorrélations		Partial Autocorrélations		ac	pac	
.....	*****	*****	1	0.512	0.512
..	**	..	**	2	0.161	-0.137
.	*	.	*	3	0.149	0.173
:	:	:	:	:	:	:
:	***	:	***	11	0.064	0.146
:	**	:	**	12	0.211	0.129
:	*	:	*	13	0.081	-0.139
:	..	:	..	14	-0.112	-0.143
:	.	:	.	:	:	:
:	..	:	..	24	0.184	0.038
Q-Statistic (24 lags) 97.002		S.E. of Correlations		0.072		

le retard 1, ce qui conduit à un modèle AR(1), soit que les autocorrélations sont tronquées après le retard 1, ce qui conduit à un modèle MA(1). Ces deux modèles ont été estimés par la méthode des moindres carrés. Pour le premier (tableau 2.5), cela revient à une régression de $DTAUR_t$ en fonction de $DTAUR_{t-1}$. C'est la raison pour laquelle une seule itération est nécessaire. Pour le second modèle (tableau 2.7), la non linéarité par rapport aux paramètres nécessite donc une procédure numérique itérative. Les autocorrélations et les autocorrélations partielles des résidus sont chaque fois présentées (tableaux 2.6 et 2.8).

Tableau 2.5: Estimation d'un premier modèle sur la série DTAUR

Convergence achieved after 1 iterations

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	0.1677768	0.1721237	0.9747454	0.330
AR(1)	0.5236992	0.0627585	8.3446742	0.000
S.E. of regression		1.130067	Sum of squared resid	240.0857

Tableau 2.6: Autocorrélations des résidus du premier modèle

Autocorrélations		Partial Autocorrélations		ac	pac	
..	**	..	**	1	0.079	0.079
.	*	.	*	2	-0.170	-0.177
:	:	:	:	3	0.126	0.166
:	..	:	..	:	:	:
:	***	:	***	11	-0.010	0.039
:	**	:	**	12	0.262	0.215
:	*	:	*	13	0.065	0.037
:	..	:	..	14	-0.194	-0.143
:	.	:	.	:	:	:
:	..	:	..	24	0.169	0.086
Q-Statistic (24 lags) 44.036		S.E. of Correlations		0.072		

Tableau 2.7: Estimation d'un second modèle sur la série DTAUR

Convergence achieved after 3 iterations

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	0.1694795	0.0797280	2.1257198	0.034
MA(1)	0.6361725	0.0729548	8.7200904	0.000
S.E. of regression		1.098952	Sum of squared resid	227.0468

Tableau 2.8: Autocorrélations des résidus du second modèle

Autocorrélations		Partial Autocorrélations		ac	pac	
.	**	.	**	1	0.032	0.032
.	**	.	**	2	0.127	0.126
.	*	.	.	3	0.088	0.082
.	.	.	.	4	0.049	0.030
.
.	.	.	.	11	0.015	-0.001
.	**	.	**	12	0.184	0.184
.	*	.	*	13	0.068	0.063
**	.	***	.	14	-0.147	-0.207
.
.	**	.	*	24	0.141	0.102
Q-Statistic (24 lags)		27.932		S.E. of Correlations		0.072

C'est le second modèle qui paraît supérieur, à la fois parce que l'écart-type des innovations est plus petit et parce que les autocorrélations sont plus petites, dans l'ensemble. Il n'en reste pas moins que $\hat{\rho}(12) = 0.184$ est significatif. Puisque $\hat{\pi}(24)$ est plus petite que $\hat{\rho}(24)$, ceci suggère qu'un polynôme autorégressif de degré 12 soit introduit :

$$(1 - \Phi L^{12}) \nabla \text{TAUR}_t = C + (1 + \theta L) \varepsilon_t.$$

Ce modèle est estimé dans le tableau 2.9. Les paramètres $\Phi = 0.27$ et $\theta = 0.62$ sont significatifs, à l'exception de la constante.

Tableau 2.9: Estimation d'un troisième modèle sur la série DTAUR

Convergence achieved after 4 iterations				
VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	0.0280216	0.1086214	0.2579750	0.796
MA(1)	0.6196608	0.0758271	8.1720264	0.000
AR(12)	0.2656115	0.0632910	4.1966719	0.000
S.E. of regression	1.050940	Sum of squared resid	193.2830	

Le tableau 2.10 montre qu'il subsiste de l'autocorrélation de retard 2 (qu'on pourrait éventuellement éliminer par l'introduction d'un terme en L^2 dans le polynôme moyenne mobile) mais le test de Box-Pierce conduit à l'acceptation du modèle puisque $Q = 24.7$.

Tableau 2.10: Autocorrélations des résidus du troisième modèle

Autocorrélations		Partial Autocorrélations		ac	pac	
.	***	.	***	1	0.006	0.006
.	*	.	*	2	0.207	0.206
.	.	.	.	3	0.036	0.035
.	.	.	.	4	0.053	0.010
.
.	.	.	.	11	-0.004	-0.039
.	*	.	*	12	-0.043	-0.027
.	**	.	**	13	0.120	0.130
.	*	.	*	14	-0.173	-0.165
.
.	*	.	*	24	0.079	0.074
Q-Statistic (24 lags)		24.723		S.E. of Correlations		0.075

Le graphique des résidus est présenté dans la figure 2.11. La série paraît assez perturbée et laisse supposer la présence de données aberrantes. En outre, on peut avoir des inquiétudes sur la qualité des prévisions réalisées à la fin de la série, pour l'année 1991. En effet, il y a une chute importante des résidus à la fin de 1990. Les prévisions seront donc fortement sous-estimées.

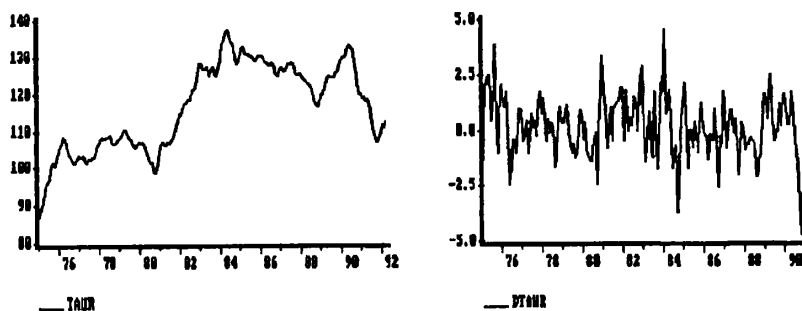


Figure 2.11: La série des résidus du troisième modèle

2.4 MODÈLES NON LINÉAIRES

Il y a de nombreuses manières d'introduire des non-linéarités dans un modèle de série temporelle. Dans cette section, nous traitons de quelques unes d'entre elles, principalement les modèles bilinéaires, les modèles avec seuil et les modèles autorégressifs conditionnellement hétéroscédastiques (« ARCH »). Ce sont ces derniers qui seront l'objet essentiel des chapitres suivants. Nous nous restreignons ici à des processus unidimensionnels.

2.4.1 Modèles bilinéaires

Les *modèles bilinéaires* (Granger et Andersen [1978]) généralisent les modèles ARMA en introduisant des termes linéaires séparément en Y et en ε , donc bilinéaires. L'équation du modèle bilinéaire général est donc

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{i=1}^P \sum_{j=1}^Q \lambda_{ij} Y_{t-i} \varepsilon_{t-j}, \quad (2.12)$$

où le processus ε est un bruit blanc gaussien (cette hypothèse peut être souvent remplacée par celle d'un bruit blanc fort avec existence de la variance). Il est désigné par le sigle $BL(p, q; P, Q)$.

L'étude de cette classe de modèles est plus complexe que celle des processus ARMA (voir Subba Rao [1981], Guegan [1987], Liu et Brockwell [1988], Liu [1989] et [1990]). Pour qu'un modèle soit pleinement utilisable du point de

vue statistique, il faut établir qu'il existe au moins une solution de l'équation qui soit un processus stationnaire et causal. En outre, pour prévoir à partir du modèle, il faut aussi que le processus soit inversible.

Des sous-classes de modèles bilinéaires ont été distinguées. Le modèle bilinéaire diagonal est défini par l'équation

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{i=1}^p \lambda_i Y_{t-i} \varepsilon_{t-i},$$

tandis que le modèle *diagonal superdiagonal* est défini par

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{i=1}^p \sum_{\substack{Q \\ j=1 \\ i \geq j}} \lambda_{ij} Y_{t-i} \varepsilon_{t-j},$$

et que le modèle *sous-diagonal* l'est par

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{i=1}^p \sum_{\substack{Q \\ j=1 \\ i < j}} \lambda_{ij} Y_{t-i} \varepsilon_{t-j}.$$

Considérons à titre d'exemple le modèle BL (0, 0; 2, 1) superdiagonal

$$Y_t = \varepsilon_t + \lambda Y_{t-2} \varepsilon_{t-1}. \quad (2.13)$$

Le processus Y est de moyenne nulle étant donné que

$$E(Y_t) = E(\varepsilon_t) + \lambda E(Y_{t-2} \varepsilon_{t-1}) = 0$$

car ε_{t-1} est indépendant du passé du processus, et en particulier de Y_{t-2} , et $E(\varepsilon_t) = 0$ pour tout t . L'autocovariance de retard h vaut

$$E(Y_t Y_{t-h}) = E[\varepsilon_t \varepsilon_{t-h} + \lambda^2 Y_{t-2} \varepsilon_{t-1} Y_{t-h-2} \varepsilon_{t-h-1} + \lambda \varepsilon_t Y_{t-h-2} \varepsilon_{t-h-1} + \lambda \varepsilon_{t-h} Y_{t-2} \varepsilon_{t-1}].$$

Si $h > 1$, nous avons

$$\begin{aligned} E(Y_t Y_{t-h}) &= E(\varepsilon_t) E(\varepsilon_{t-h}) + \lambda^2 E(\varepsilon_{t-1}) E(Y_{t-2} Y_{t-h-2} \varepsilon_{t-h-1}) \\ &\quad + \lambda E(\varepsilon_t) E(Y_{t-h-2} \varepsilon_{t-h-1}) + \lambda E(Y_{t-2} \varepsilon_{t-h}) E(\varepsilon_{t-1}) \\ &= 0. \end{aligned}$$

Pour $h = 1$, le regroupement des facteurs est différent pour le dernier terme mais conduit au même résultat :

$$\begin{aligned} E(Y_t Y_{t-1}) &= E(\varepsilon_t) E(\varepsilon_{t-1}) + \lambda^2 E(\varepsilon_{t-1}) E(Y_{t-2} Y_{t-3} \varepsilon_{t-2}) \\ &\quad + \lambda E(\varepsilon_t) E(Y_{t-3} \varepsilon_{t-2}) + \lambda E(Y_{t-2}) E(\varepsilon_{t-1}^2) \\ &= 0. \end{aligned}$$

Enfin, en remplaçant h par 0, nous trouvons une relation de récurrence pour la variance marginale $E(Y_t^2)$:

$$E(Y_t^2) = E(\varepsilon_t^2) + \lambda^2 E(Y_{t-2}^2) E(\varepsilon_{t-1}^2) + 2\lambda E(\varepsilon_t) E(Y_{t-2}) E(\varepsilon_{t-1})$$

ou, comme le dernier terme s'annule,

$$E(Y_t^2) = \sigma^2 + \lambda^2 \sigma^2 E(Y_{t-2}^2).$$

Il existe une solution stationnaire du second ordre de l'équation (2.13) à condition que $\lambda^2 \sigma^2 < 1$. La variance marginale vaut alors

$$V(Y_t) = \frac{\sigma^2}{1 - \lambda^2 \sigma^2}.$$

On peut montrer (Guegan [1981]) que le modèle a une solution stationnaire causale unique, donnée par

$$Y_t = \sum_{j=1}^{\infty} \lambda^j \varepsilon_{t-2j} \prod_{k=1}^j \varepsilon_{t-2k+1} + \varepsilon_t.$$

Le processus se comporte comme un processus bruit blanc, du point de vue de sa fonction d'autocorrélation. Il est pourtant prévisible. La prévision en t d'horizon 1 vaut $\lambda Y_{t-1} \varepsilon_t$ et la proportion de la variance qui est expliquée, $\text{var}(\lambda Y_{t-1} \varepsilon_t) / \text{var}(Y_t)$, vaut $\lambda^2 \sigma^2$. Par ailleurs, la variance conditionnelle de Y_t étant donné Y_{t-2} vaut

$$\begin{aligned} E(Y_t^2 | Y_{t-2}) &= E[\varepsilon_t^2 + \lambda^2 Y_{t-2}^2 \varepsilon_{t-1}^2 + 2\lambda \varepsilon_t Y_{t-2} \varepsilon_{t-1} | Y_{t-2}] \\ &= \sigma^2 + \lambda^2 \sigma^2 Y_{t-2}^2 \end{aligned}$$

ce qui montre que le modèle est conditionnellement hétéroscédastique.

La figure 2.12 montre une réalisation d'un processus bilinéaire avec $\lambda = 0.6$ et $\sigma = 1$, et une moyenne égale à 5. La série a été générée au moyen des équations suivantes :

$$Y_t = \begin{cases} \varepsilon_t & \text{si } t = 1, 2 \\ 5 + 0.6\varepsilon_{t-1}(Y_{t-2} - 5) + \varepsilon_t & \text{si } t \geq 3. \end{cases}$$

La figure 2.13 montre les autocorrélations de la série des 100 dernières valeurs qui ressemblent assez à celles d'un bruit blanc, comme on pouvait s'y attendre. Elle montre aussi les autocorrélations de la série des Y_t^2 , dont plusieurs sont significatives. On comprend mieux que la proportion de la variance expliquée doit valoir ici 0.36 quand on regarde la figure 2.14 qui montre Y_t en fonction de Y_{t-1} .

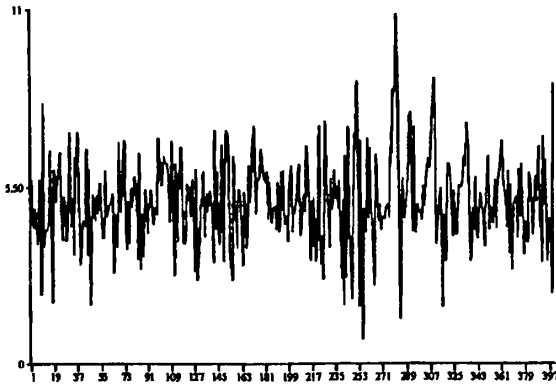


Figure 2.12: Série artificielle générée par un processus bilinéaire

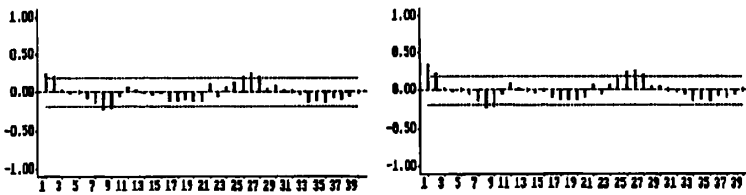


Figure 2.13

Autocorrélations de la série artificielle de longueur 100 générée par un processus bilinéaire et des carrés de cette série

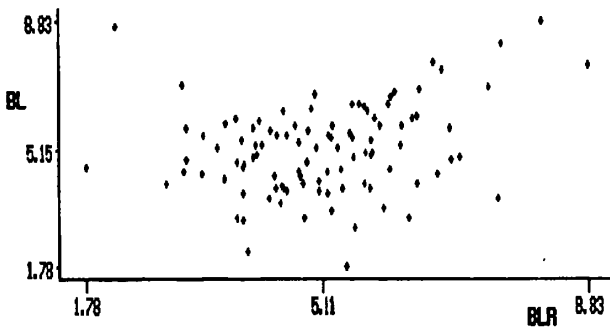


Figure 2.14

La série artificielle générée par un processus bilinéaire en fonction de la série retardée d'une unité

Remarques

1. La manière de paramétrer la constante est différente de celle employée le plus souvent. Cette paramétrisation permet de ne pas modifier les conditions de stationnarité.
2. Les moments de Y d'ordre impair sont nuls mais les moments d'ordre pair ne sont pas tous finis. Ils ne le sont pas si le support de la loi de ε est infini, ce qui est le cas notamment pour la loi normale. Par ailleurs, on peut montrer que la condition d'inversibilité du processus est $2\lambda^2\sigma^2 < 1$ (Guegan [1981]).
3. Il faut être prudent en matière de génération de données artificielles selon un modèle non linéaire. En effet, les algorithmes pour générer les nombres pseudo-aléatoires sont déterministes, typiquement non linéaires. Il se pourrait qu'un choix malheureux ait une incidence sur des résultats d'une étude empirique par simulation. Il est donc important de préciser les conditions de cette étude, voir par exemple Ashley *et al.* [1986].

2.4.2 Modèles autorégressifs à seuils

Les modèles *autorégressifs à seuils* (« threshold autoregressive » ou TAR) ont été proposés par Tong [1978] (voir aussi Tong [1983], [1990]). Tong les a introduit comme des approximations discrètes de modèles non linéaires en temps continu sous forme d'équations différentielles non linéaires. Ils permettent de reproduire des phénomènes tels qu'un cycle limite et une dépendance entre amplitude et fréquence. Supposons que le processus Y vérifie au temps t une équation parmi plusieurs équations différentes selon la valeur d'une variable (autre que Y_t). Chaque équation correspond à un *régime*. Dans le cas d'un seuil unique et d'une variable X_t , on aura par exemple :

$$Y_t = \begin{cases} \sum_{i=1}^{p_1} \phi^{(1)} Y_{t-i} + \varepsilon_t & \text{si } X_t \leq \alpha \\ \sum_{i=1}^{p_2} \phi^{(2)} Y_{t-i} + \varepsilon_t & \text{si } X_t > \alpha. \end{cases} \tag{2.14}$$

Tong a considéré l'existence de plusieurs seuils. La variable X_t , est soit une variable exogène, soit une valeur retardée de δ de la variable étudiée, $Y_{t-\delta}$. Dans ce dernier cas, on parle de modèle SETAR (« self excited threshold AR »). Une constante peut être ajoutée à chacun des sous-modèles autorégressifs. D'autres généralisations sont possibles. On peut employer des sous-modèles ARMA au lieu de AR et permettre aux variances des ε_t d'être différentes entre les sous-modèles (Mélard et Roy [1988]). Chan et Tong [1986] ont introduit la possibilité de changement lisse de modèle (« smooth threshold autoregression » ou STAR).

Un cas particulier simple est le suivant :

$$Y_t = \begin{cases} \phi^{(1)}Y_{t-1} + \varepsilon_t & \text{si } Y_{t-1} \leq \alpha \\ \phi^{(2)}Y_{t-1} + \varepsilon_t & \text{si } Y_{t-1} > \alpha. \end{cases} \quad (2.15)$$

Une condition nécessaire et suffisante d'existence d'une solution stationnaire et ergodique est $\phi^{(1)} < 1$, $\phi^{(2)} < 1$ et $\phi^{(1)}\phi^{(2)} < 1$. Même si la distribution de ε est symétrique par rapport à zéro, la distribution marginale n'est pas symétrique, sauf si $\phi^{(1)} = \phi^{(2)}$ (Petrucci et Woolford [1984] ; pour une étude de la distribution marginale des Y_t , dans le cas d'un processus ε gaussien, voir Andel *et al.* [1984]).

Pour illustrer les modèles avec seuil, nous avons choisi l'exemple suivant, d'équation :

$$Y_t = \begin{cases} 5 - 0.2(Y_{t-1} - 5) + \varepsilon_t & \text{si } Y_{t-1} \leq 6.5 \\ 5 + 0.9(Y_{t-1} - 5) + \varepsilon_t & \text{si } Y_{t-1} > 6.5. \end{cases}$$

Dans la figure 2.15 est représentée une série générée par ce processus, en employant la série de la figure 2.1 comme réalisation du bruit blanc.

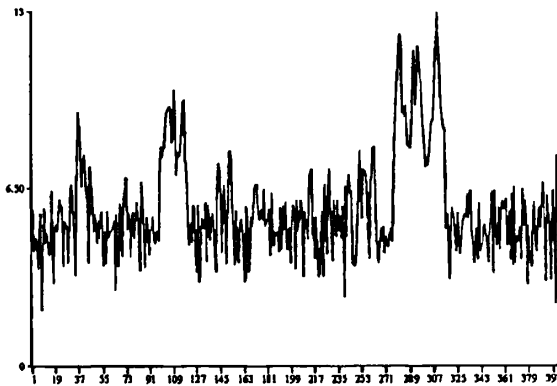


Figure 2.15: Série artificielle générée par un processus autorégressif à seuil

La figure 2.16 montre les autocorrélations de la série des 100 dernières valeurs et la série Y_t , en fonction de Y_{t-1} . On voit clairement la non linéarité de la relation entre Y_t et Y_{t-1} .

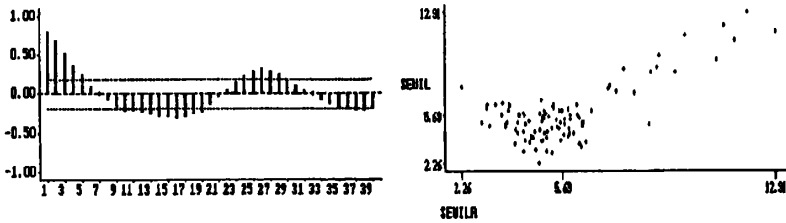


Figure 2.16

Autocorrélations de la série artificielle de longueur 100 générée par un processus autorégressif à seuil (à gauche) et la série en fonction de la série retardée d'une unité (à droite)

2.4.3 Modèles autorégressifs conditionnellement hétéroscédastiques

La classe des modèles *autorégressifs conditionnellement hétéroscédastiques* («autoregressive conditional heteroscedastic» ou ARCH) a été proposée par Engle en 1982 (cfr. chapitre 1). Le but est de permettre à la variance du bruit de varier conditionnellement au passé de la série. Le modèle ARCH(q) est défini par les équations

$$\begin{cases} \varepsilon_t = v_t \sqrt{h_t} \\ h_t = c + \sum_{i=1}^q a_i \varepsilon_{t-i}^2 \\ v_t \sim \text{i.i.d. } N(0, 1). \end{cases} \quad (2.16)$$

Il apparaît comme un bruit blanc gaussien, multiplié à chaque temps t par une variable aléatoire dont le carré dépend, de manière linéaire, des q dernières valeurs du processus.

On peut écrire le modèle sous la forme

$$\begin{cases} \varepsilon_t | \underline{\varepsilon}_{t-1} \sim N(0, h_t) \\ h_t = c + \sum_{i=1}^q a_i \varepsilon_{t-i}^2, \end{cases} \quad (2.17)$$

ce qui montre que la distribution conditionnelle de ε_t , est normale centrée et de variance h_t , laquelle dépend linéairement des q dernières valeurs du processus. Une troisième formulation fait apparaître ε comme une différence de martingale, puisque

$$E(\varepsilon_t | \underline{\varepsilon}_{t-1}) = 0$$

mais hétéroscédastique puisque

$$V(\varepsilon_t | \underline{\varepsilon}_{t-1}) = c + \sum_{i=1}^q a_i \varepsilon_{t-i}^2. \quad (2.18)$$

Enfin, en notant $u_t = \varepsilon_t^2 - h_t$, il vient une quatrième formulation :

$$\varepsilon_t^2 = c + \sum_{i=1}^q a_i \varepsilon_{t-i}^2 + u_t, \quad (2.19)$$

ce qui montre une forme AR(q) pour les ε_t^2 , avec toutefois la remarque que les variables $u_t = v_t^2 h_t - h_t = (v_t^2 - 1)h$ ne sont pas de variance constante mais sont de moyenne nulle et non corrélées entre elles. Notons que les deux dernières formulations donnent lieu à des définitions plus générales de modèles ARCH où la normalité n'est pas requise.

Le modèle ARCH général sera étudié dans le chapitre 3 ainsi que les généralisations telles que les modèles GARCH (Bollerslev [1986]), modèle de régression à erreurs ARCH (Engle [1982]), ARMA à erreurs ARCH (Weiss [1984]). Nous nous limitons ici à une étude préliminaire du modèle ARCH(1) et du modèle AR(1) à erreurs ARCH(1).

Considérons donc le modèle ARCH(1). Le processus de différence de martingale ε est ARCH(1) si et seulement si

$$\varepsilon_t = v_t \sqrt{c + a\varepsilon_{t-1}^2}, \quad v_t \sim \text{i.i.d. } N(0, 1). \quad (2.20)$$

Il faut évidemment que $a \geq 0$ et que $c > 0$, sans quoi la variance conditionnelle pourrait être négative. Pour que le processus soit stationnaire du second ordre, la variance marginale de ε doit être constante. Or elle vérifie l'équation

$$V(\varepsilon_t) = c + aV(\varepsilon_{t-1})$$

d'où l'on conclut que $a < 1$ et que

$$V(\varepsilon_t) = \frac{c}{1-a}. \quad (2.21)$$

Soit h un entier strictement positif. Remarquons d'abord qu'en vertu du théorème des projections itérées (équation (2.2)), et compte tenu du fait que $\underline{\varepsilon_{t-h}} \subset \underline{\varepsilon_{t-1}}$, on a

$$\begin{aligned} E(\varepsilon_t \mid \underline{\varepsilon_{t-h}}) &= E\left(E(\varepsilon_t \mid \underline{\varepsilon_{t-1}}) \mid \underline{\varepsilon_{t-h}}\right) \\ &= E\left(0 \mid \underline{\varepsilon_{t-h}}\right) \\ &= 0. \end{aligned}$$

Il est intéressant de comparer la variance marginale à la variance conditionnelle $V(\varepsilon_t \mid \underline{\varepsilon_{t-h}})$. Par remplacements successifs, on déduit de (2.19) :

$$\varepsilon_t^2 = c(1 + a + \dots + a^{h-1}) + a^h \varepsilon_{t-h}^2 + u_t + a u_{t-1} + \dots + a^{h-1} u_{t-h+1}.$$

Prenant l'espérance conditionnelle des deux membres, conditionnellement à l'information $\underline{\varepsilon}_{t-h}$, on obtient

$$\begin{aligned} V(\varepsilon_t | \underline{\varepsilon}_{t-h}) &= c(1 + a + \dots + a^{h-1}) + a^h \varepsilon_{t-h}^2 \\ &= c \frac{1 - a^h}{1 - a} + a^h \varepsilon_{t-h}^2. \end{aligned} \tag{2.22}$$

Lorsque h tend vers l'infini, la variance conditionnelle converge vers la variance marginale donnée par (2.21). La différence entre la variance conditionnelle et la variance marginale peut s'écrire

$$V(\varepsilon_t | \underline{\varepsilon}_{t-h}) - V(\varepsilon_t) = a^h [\varepsilon_{t-h}^2 - V(\varepsilon_{t-h})].$$

C'est une fonction très simple de la différence entre le carré de l'innovation et sa moyenne. Au plus l'innovation est grande en valeur absolue, au plus la variance conditionnelle future est au-dessus de la variance marginale.

On peut vérifier (Gourieroux, [1992]) que les conditions imposées au processus ε sont compatibles. Notons que les résultats ci-dessus ne dépendent pas de l'hypothèse de normalité, c'est-à-dire du caractère gaussien du processus v .

Pour déterminer les conditions de stationarité et pour trouver la distribution marginale du processus, il faut procéder par récurrence. Illustrons ceci sur le moment d'ordre 4, dans le cas gaussien. A partir de (2.20), on peut écrire

$$\begin{aligned} E(\varepsilon_t^4 | \underline{\varepsilon}_{t-1}) &= E(v_t^4 h_t^2 | \underline{\varepsilon}_{t-1}) \\ &= E(v_t^4 | \underline{\varepsilon}_{t-1}) E(h_t^2 | \underline{\varepsilon}_{t-1}) \\ &= 3(c + a\varepsilon_{t-1}^2)^2, \end{aligned}$$

compte tenu de ce que le moment d'ordre 4 d'une loi normale $N(0, 1)$ vaut 3. En prenant l'espérance marginale par rapport à $\underline{\varepsilon}_{t-1}$, il vient

$$\begin{aligned} E(\varepsilon_t^4) &= 3E(c^2 + 2ac\varepsilon_{t-1}^2 + a^2\varepsilon_{t-1}^4) \\ &= 3[c^2 + 2acE(\varepsilon_{t-1}^2) + a^2E(\varepsilon_{t-1}^4)] \\ &= 3\left[c^2 + \frac{2ac^2}{1-a} + a^2E(\varepsilon_{t-1}^4)\right] \end{aligned}$$

d'où la stationnarité implique

$$E(\varepsilon_t^4) = \frac{3c^2(1+a)}{(1-3a^2)(1-a)}.$$

La condition d'existence des moments d'ordre 4 est $3a^2 < 1$. On en déduit que le coefficient d'aplatissement de Fisher (« kurtosis ») associé à la distribution marginale de ε vaut

$$\kappa = \frac{E(\varepsilon^4)}{E^2(\varepsilon^2)} - 3 = 3 \frac{1 - a^2}{1 - 3a^2} - 3 > 0,$$

cette dernière étant la valeur de κ pour une loi normale. Dans ce texte, on utilise une mesure différente mais équivalente d'aplatissement

$$k = \frac{E(\varepsilon^4)}{E^2(\varepsilon^2)} = 3 \frac{1 - a^2}{1 - 3a^2} > 3,$$

cette dernière étant la valeur de k pour une loi normale. Les queues de la distribution marginale d'un processus ARCH(1) sont donc plus épaisses que pour un processus gaussien. On dit que cette distribution est leptokurtique.

Plus généralement, les moments centrés d'ordre impair, s'ils existent, sont nuls, par symétrie. En supposant que le processus démarre infiniment loin dans le passé avec les $2r$ premiers moments finis, le moment d'ordre $2r$ existe si et seulement si (Engle [1982])

$$a^r \prod_{i=1}^r (2i - 1) < 1.$$

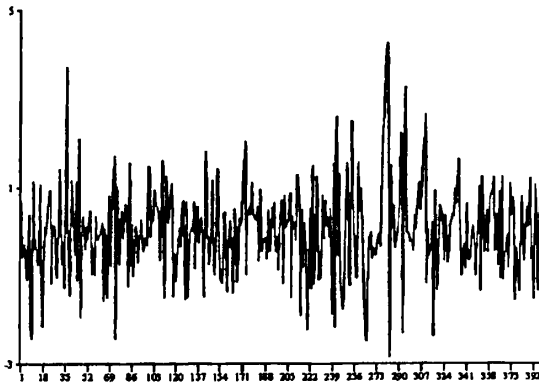


Figure 2.17: Série artificielle générée par un processus ARCH(1)

La figure 2.17 montre une série artificielle générée par le modèle ARCH(1) d'équation :

$$\varepsilon_t = v_t \sqrt{0.5 + 0.5\varepsilon_{t-1}^2}, \quad v_t \sim \text{i.i.d. } N(0, 1)$$

Puisque $a = 0.5$, les moments marginaux d'ordre 2 et 4 existent et valent respectivement 1 et 9, d'où $\kappa = 6$. La figure 2.18 présente la fonction d'autocorrélation de la série des 100 dernières données et de la série des carrés de ces données.

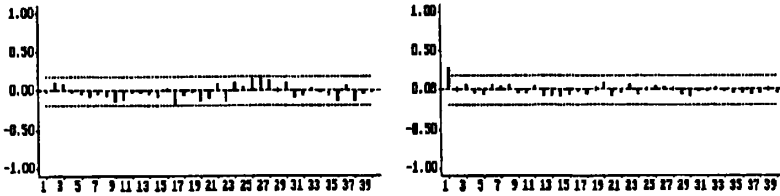


Figure 2.18

Autocorrélations de la série artificielle de longueur 100 générée par un processus ARCH(1) (à gauche) et de la série des carrés (à droite)

Nous envisageons maintenant un processus AR(1) à erreurs ARCH(1), défini par l'équation

$$\begin{cases} Y_t = \mu + \phi Y_{t-1} + \varepsilon_t, & |\phi| < 1 \\ E(\varepsilon_t | \varepsilon_{t-1}) = 0 \\ V(\varepsilon_t | \varepsilon_{t-1}) = c + a\varepsilon_{t-1}^2. \end{cases} \quad (2.23)$$

On a, pour l'espérance conditionnelle, l'expression suivante

$$E(Y_t | Y_{t-h}) = \mu + \phi E(Y_{t-1} | Y_{t-h})$$

ce qui montre que les prévisions non linéaires de Y s'obtiennent comme les prévisions linéaires d'un modèle AR(1). Plus généralement

$$Y_t = \mu \frac{1 - \phi^h}{1 - \phi} + \phi^h Y_{t-h} + \varepsilon_t + \phi \varepsilon_{t-1} + \dots + \phi^{h-1} \varepsilon_{t-h+1}. \quad (2.24)$$

En prenant l'espérance conditionnelle de (2.24), on obtient

$$E(Y_t | Y_{t-h}) = \mu \frac{1 - \phi^h}{1 - \phi} + \phi^h Y_{t-h}.$$

La variance conditionnelle s'obtient comme suit, compte tenu de (2.22) :

$$\begin{aligned} V(Y_t | Y_{t-h}) &= V\left(\mu \frac{1 - \phi^h}{1 - \phi} + \phi^h Y_{t-h} + \varepsilon_t + \phi \varepsilon_{t-1} + \dots + \phi^{h-1} \varepsilon_{t-h+1} \mid Y_{t-h}\right) \\ &= V(\varepsilon_t | \varepsilon_{t-h}) + \phi^2 V(\varepsilon_{t-1} | \varepsilon_{t-h}) + \dots + \phi^{2(h-1)} V(\varepsilon_{t-h+1} | \varepsilon_{t-h}) \\ &= \sum_{j=0}^{h-1} \phi^{2j} \left(c \frac{1 - a^{h-j}}{1 - a} + a^{h-j} \varepsilon_{t-h}^2 \right) \end{aligned}$$

$$= \frac{c}{1-a} \left(\frac{1-\phi^{2h}}{1-\phi^2} - a \frac{a^h - \phi^{2h}}{a - \phi^2} \right) + a \frac{a^h - \phi^{2h}}{a - \phi^2} \varepsilon_{t-h}^2.$$

Elle n'est pas constante puisqu'elle dépend de ε_{t-h}^2 . On en déduit que les intervalles de prévision, d'horizon h fixé, d'un modèle AR(1) avec erreurs ARCH(1) seront d'autant plus larges que la variabilité est grande, contrairement au modèle AR(1) où les intervalles de prévision dépendent de h mais pas de t .

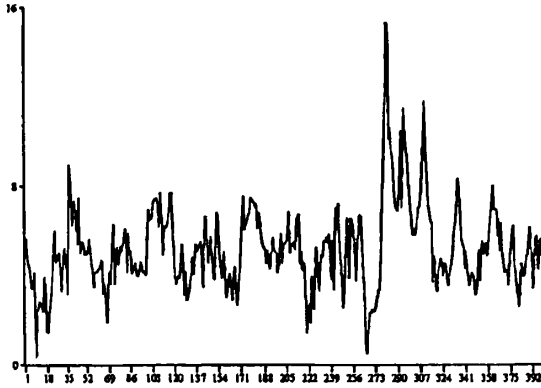


Figure 2.19

Série artificielle générée par un processus AR(1) à erreurs ARCH(1)

La figure 2.19 montre une série artificielle générée par le modèle AR(1) à erreurs ARCH(1), défini par :

$$\begin{cases} Y_t = 5 + 0.8(Y_{t-1} - 5) + \varepsilon_t \\ \varepsilon_t = v_t \sqrt{0.5 + 0.5\varepsilon_{t-1}^2} \\ v_t \sim \text{i.i.d. } N(0, 1). \end{cases}$$

On constate dans la figure (2.20) que les premières autocorrélations de la série des 100 dernières données ressemblent à celle d'un processus AR(1).

2.4.4 Quelques autres modèles non linéaires

Le contexte général des modèles non linéaires est le *développement de Volterra* qui généralise, dans un certain sens, la décomposition de Wold. Ce développement est de la forme

$$Y_t = \sum_{i=0}^{\infty} \theta_i \varepsilon_{t-i} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \theta_{ij} \varepsilon_{t-i} \varepsilon_{t-j} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \theta_{ijk} \varepsilon_{t-i} \varepsilon_{t-j} \varepsilon_{t-k} + \dots \quad (2.25)$$

où ε est un bruit blanc fort.

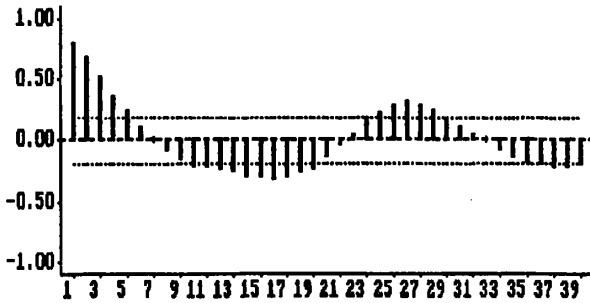


Figure 2.20

Autocorrélation de la série artificielle de longueur 100 générée par un processus AR(1) à erreurs ARCH(1)

Une autre approche très générale est celle des *modèles dépendant de l'état* (« state-dependent model » ou SDM) de Priestley [1980]. Considérons un modèle d'ordre p . Appelons $\underline{y}_{t-1} = (Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})$ le vecteur d'état. Supposons pour Y_t un modèle autorégressif d'ordre p dont les coefficients $\phi_i(\underline{y}_{t-1})$, $i = 1, \dots, p$ sont fonction du vecteur d'état :

$$Y_t = \sum_{i=1}^p \phi_i(\underline{y}_{t-1}) Y_{t-i} + \varepsilon_t. \tag{2.26}$$

La généralisation aux modèles ARMA(p, q) est immédiate, en posant $\underline{y}_{t-1} = (Y_{t-1}, \dots, Y_{t-p}, \varepsilon_{t-1}, \dots, \varepsilon_{t-q})$:

$$Y_t = \sum_{i=1}^p \phi_i(\underline{y}_{t-1}) Y_{t-i} + \varepsilon_t - \sum_{j=1}^q \theta_j(\underline{y}_{t-1}) \varepsilon_{t-j}. \tag{2.27}$$

Cette classe de modèles contient en particulier :

- (a) les modèles bilinéaires, si les ϕ_i sont constants et les $\theta_j(\underline{y}_{t-1})$ sont des fonctions linéaires de \underline{y}_{t-1} ;
- (b) les modèles à seuils SETAR, si $\phi_i(\underline{y}_{t-1}) = \phi_i^{(1)}$, $i = 1, \dots, p$, si $Y_{t-\delta} \leq \alpha$ et $\phi_i^{(2)}$, dans le cas contraire ;
- (c) les modèles *autorégressifs exponentiels* de Haggan et Ozaki [1981], en posant

$$\phi_i(\underline{y}_{t-1}) = a_i + b_i e^{-Y_{t-1}^2/c}.$$

Parmi les autres approches, celle de Lawrance et Lewis [1985] mérite d'être mentionnée. Le modèle est défini, en probabilité, comme suit

$$Y_t = \begin{cases} \phi^{(1)}Y_{t-1} + \varepsilon_t & \text{avec la probabilité } p_1 \\ \phi^{(2)}Y_{t-1} + \varepsilon_t & \text{avec la probabilité } p_2 \\ \varepsilon_t & \text{avec la probabilité } 1 - p_1 - p_2, \end{cases} \quad (2.28)$$

où $p_1 > 0$, $p_2 > 0$, $p_1 + p_2 < 1$, $0 < \phi_1, \phi_2 < 1$, ε est un processus de type bruit blanc fort choisi de manière adéquate pour que la distribution marginale de Y soit exponentielle de moyenne 1. On peut alors montrer que la loi de ε est un mélange de trois exponentielles de paramètres spécifiés.

2.5 TESTS DE LINÉARITÉ

2.5.1 Tests basés sur les autocorrélations

Il convient avant tout de rappeler la propriété suivante :

Si ε est un bruit blanc fort, $g(\varepsilon_t)$ est non corrélé avec $g(\varepsilon_{t+h})$, pour tout $h \neq 0$, pour toute fonction g .

Les autocorrélations des carrés d'une série sont utiles pour mettre en évidence des non linéarités (Granger et Andersen, [1978]). Nous avons déjà constaté qu'une série peut sembler non autocorrélée alors que son carré présente une autocorrélation marquée, illustrant de ce fait

- (a) que la non corrélation n'entraîne pas l'indépendance ;
- (b) qu'une série peut être non prédictible par un modèle linéaire mais prédictible par le modèle non linéaire adéquat.

Granger et Newbold [1976] ont montré que si Y est un processus gaussien dont les autocorrélations sont notées $\rho_Y(h)$, les autocorrélations du processus Y^2 sont $\rho_{Y^2}(h)$. On peut estimer les autocorrélations de Y^2 à partir d'une réalisation de Y au moyen de

$$\hat{\rho}_{Y^2}(h) = \frac{\sum_{t=h+1}^T (Y_t^2 - \hat{\sigma}^2)(Y_{t-h}^2 - \hat{\sigma}^2)}{\sum_{t=1}^T (Y_t^2 - \hat{\sigma}^2)^2},$$

où $\hat{\sigma}^2$ est l'estimation de la variance de Y . Si on représente graphiquement ces $\hat{\rho}_{Y^2}(h)$ en fonction des carrés des $\hat{\rho}_Y(h)$, pour un processus gaussien on devrait trouver les points autour de la première bissectrice. On peut appliquer cette technique aux résidus d'un modèle linéaire ou non linéaire.

Par ailleurs, McLeod et Li [1983] ont étudié la distribution asymptotique des autocorrélations des carrés des résidus d'un modèle ARMA(p, q) à erreurs

indépendantes. Ils ont montré que la loi marginale est une loi normale de moyenne 0 et de variance $1/T$ et qu'il y a non corrélation asymptotique. De ce fait, la statistique de type Box-Pierce (2.10) correspondante a une loi χ_H^2 , où H est le nombre d'autocorrélations utilisées.

2.5.2 Tests basés sur la régression

Keenan [1985] a proposé un test d'absence de termes quadratiques dans un développement de Volterra d'ordre 2. La méthode consiste en les étapes suivantes :

- 1) régression de Y_t sur $\{1, Y_{t-1}, \dots, Y_{t-m}\}$ qui donne des résidus $\hat{\varepsilon}_t$, $t = m + 1, \dots, T$;
- 2) régression de Y_t^2 sur $\{1, Y_{t-1}, \dots, Y_{t-m}\}$ qui donne des résidus $\hat{\eta}_t$, $t = m + 1, \dots, T$;
- 3) régression de $\hat{\varepsilon}_t$ sur $\hat{\eta}_t$, $t = m + 1, \dots, T$ (sans constante), et test de la nullité du coefficient de régression, par un test de Student à $T - 2m - 1$ degrés de liberté (le nombre de résidus diminué du nombre de paramètres estimés).

La méthode des *multiplicateurs de Lagrange* a été employée pour obtenir une classe de tests pour détecter la non linéarité. Leur puissance a été étudiée dans le contexte des modèles bilinéaires dans le cas particulier où les cumulants d'ordre 3 et 4 du bruit blanc sont supposés nuls (Guegan et Pham, [1992]).

Plus généralement, Tsay [1991] propose et étudie une procédure qui combine les tests de multiplicateurs de Lagrange et les *autorégressions ordonnées*. Les premiers consistent à ajouter des variables explicatives dûment choisies à un modèle autorégressif, transformant celui-ci en modèle de régression à erreurs AR. Pour détecter la possibilité de bilinéarité, on choisit des variables comme $Y_{t-i}\hat{\varepsilon}_{t-i}$ ou $\hat{\varepsilon}_{t-i}\hat{\varepsilon}_{t-i-1}$, où les $\hat{\varepsilon}_t$ sont les résidus du modèle autorégressif. Pour un modèle autorégressif exponentiel, il propose $Y_{t-i} \exp\{-Y_{t-1}^2/c\}$, où c est une constante de normalisation, par exemple $c = \max\{|Y_{t-1}|\}$. Pour un modèle à seuil, on peut employer une variable indicatrice $\mathbf{1}_{Y_{t-1} > \alpha} = 1$, si $Y_{t-1} > \alpha$, et 0, dans le cas contraire.

Les autorégressions ordonnées sont employées pour détecter le besoin d'un modèle à seuil de type SETAR. Pour un délai donné δ et un ordre m du modèle autorégressif, on ordonne les temps t en fonction croissante de $Y_{t-\delta}$ et on estime le modèle par régression linéaire appliquée par récurrence, en déterminant successivement les résidus récursifs normalisés. Cette méthode améliore la méthode CUSUM de Petrucci et Davies [1986]. Au lieu d'employer toutes les données, il est préférable d'employer une fenêtre afin de réaliser un ajustement « local ». Si la fenêtre est de largeur L , seules L données sont utilisées ; on ajoute donc chaque fois un point dans la récurrence mais, si le nombre de points utilisés dépasse L , on retire aussi chaque fois un point. De cette manière, le franchissement du seuil apparaît comme la nécessité de changer de modèle.

2.5.3 Autres tests

Des méthodes d'analyse spectrale ont aussi été utilisées. Subba Rao et Gabr [1980] ont employé le *bispectre*, $f_2(\omega_1, \omega_2)$, qui est la transformée de Fourier du moment d'ordre 3, défini comme suit

$$\gamma(h, k) = E[(Y_t - m_t)(Y_{t+h} - m_{t+h})(Y_{t+k} - m_{t+k})]$$

Pour un processus ayant une décomposition de Wold en termes d'un bruit blanc fort ε dont le moment d'ordre 3, μ_3 , est fini, le rapport entre le carré du module du bispectre et le produit $f(\omega_1)f(\omega_2)f(\omega_1 + \omega_2)$ est constant et proportionnel à μ_3^2 . La comparaison de l'estimation de ce produit avec l'estimation du bispectre fournit un test de linéarité et un test de normalité (car le bispectre est nul pour un processus gaussien). Hinich [1982] a amélioré le test de normalité et a rendu le test de linéarité plus robuste. Ashley *et al.* [1986] ont employé ces tests par rapport à différents types de modèles non linéaires.

Pour certaines classes de modèles bilinéaires, mais sans spécification de la densité du bruit blanc, Benghabrit et Hallin [1992] ont amélioré cette théorie et ont montré que les statistiques de tests localement asymptotiquement optimaux sont des coefficients d'autocorrélation cubique. Il en ont également donné la version non paramétrique, qui se ramène à des coefficients d'autocorrélation cubique basés sur les rangs, ou sur les rangs signés, dans le cas d'une densité symétrique. Enfin, ils ont calculé l'efficacité asymptotique de ces tests non paramétriques par rapport aux tests de multiplicateurs de Lagrange.

CHAPITRE 3

LES MODÈLES HÉTÉROSCÉDASTIQUES UNIVARIÉS

par

Christian GOURIEROUX

3.1 DIVERSES MODÉLISATIONS

Des formalisations plus ou moins simples à utiliser, plus ou moins adaptées à l'analyse des séries financières ont été introduites dans la littérature pour prendre en compte les phénomènes d'hétéroscédasticité conditionnelle. Les premiers modèles proposés cherchaient essentiellement à introduire des bruits blancs conditionnellement hétéroscédastiques ; cette démarche s'explique par l'importance de l'hypothèse de marche aléatoire en finance, hypothèse qui conduit à modéliser directement les rendements, c'est-à-dire les accroissements relatifs de prix, sous forme de bruit blanc.

3.1.1 Modèles ARCH(q) (Engle [1982])

L'aspect hétéroscédastique est introduit par l'intermédiaire d'une dynamique autorégressive sur les carrés du bruit. Le modèle est défini par :

$$\left\{ \begin{array}{l} Y_t = \varepsilon_t, \\ \text{avec : } E(\varepsilon_t / \varepsilon_{t-1}) = 0, \\ E(\varepsilon_t^2 / \varepsilon_{t-1}) = h_t = c + \sum_{i=1}^q a_i \varepsilon_{t-i}^2, a_q \neq 0. \end{array} \right. \quad (3.1)$$

La volatilité de la date t est donc fonction des carrés des écarts à la moyenne observés dans le passé proche. Si les coefficients a_i , $i = 1, \dots, q$, sont tous positifs et assez grands, il y a une certaine persistance des niveaux de volatilité : on observe alors des périodes de fortes volatilités suivies de périodes de faibles volatilités. Si le polynôme caractéristique $1 - \sum_{i=1}^q a_i z^i$ admet des racines complexes, on peut observer des phénomènes cycliques sur les niveaux de volatilité.

3.1.2 Modèles GARCH (p, q) (Bollerslev [1986])

La formulation précédente fondée sur une relation auto-régressive peut en pratique conduire à un ordre de retards q assez grand, donc à l'estimation d'un nombre important de paramètres. Par analogie avec la démarche usuelle de Box-Jenkins, Bollerslev proposa d'introduire une dynamique autorégressive moyenne mobile et considéra le modèle suivant :

$$\begin{cases} Y_t = \varepsilon_t, \\ \text{avec } E(\varepsilon_t / \varepsilon_{t-1}) = 0, \\ E(\varepsilon_t^2 / \varepsilon_{t-1}) = h_t = c + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}, \alpha_q \neq 0, \beta_p \neq 0. \end{cases} \quad (3.2)$$

Pour bien mettre en évidence l'interprétation ARMA, il est nécessaire de faire apparaître l'innovation relative au carré ε_t^2 , c'est-à-dire $u_t = \varepsilon_t^2 - E(\varepsilon_t^2 / \varepsilon_{t-1}) = \varepsilon_t^2 - h_t$. Remarquons que cette innovation est en général conditionnellement hétéroscédastique ; ainsi si la loi conditionnelle de ε_t sachant ε_{t-1} est normale, nous avons :

$$\begin{aligned} E(u_t^2 / \varepsilon_{t-1}) &= E(\varepsilon_t^4 / \varepsilon_{t-1}) - (E(\varepsilon_t^2 / \varepsilon_{t-1}))^2 \\ &= 2h_t^2. \end{aligned}$$

Introduisant l'innovation dans l'écriture dynamique, nous obtenons :

$$\begin{aligned} \varepsilon_t^2 - u_t &= c + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j (\varepsilon_{t-j}^2 - u_{t-j}), \\ \varepsilon_t^2 &= c + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) \varepsilon_{t-i}^2 + u_t - \sum_{j=1}^p \beta_j u_{t-j}, \end{aligned}$$

avec les conventions $\alpha_i = 0$, si $i > q$, $\beta_j = 0$, si $j > p$. Ainsi la dynamique correspond en général à un modèle ARMA [Max (p, q), p] sur le carré du processus.

3.1.3 Modèles ARMAX-GARCH

L'hypothèse de marche aléatoire sur les prix ou les logarithmes de prix peut cependant être remise en cause ou au moins testée à partir des données. Dans ce cas la série des rendements ne correspond plus nécessairement à un bruit blanc ; le rendement net anticipé peut dépendre de divers facteurs explicatifs ou posséder une dynamique assez régulière. Ceci conduit à introduire l'effet de variables additionnelles sur la moyenne et à placer la formulation GARCH sur le processus d'innovation correspondant. Plus précisément, on peut considérer des modèles du type :

$$\left\{ \begin{array}{l} \Phi(L)Y_t = A(L)X_t + \Theta(L)\varepsilon_t, \\ \text{où } X_t \text{ sont des variables explicatives observables} \\ \text{et où } E\left(\varepsilon_t/\varepsilon_{t-1}, X_t\right) = 0 \\ V\left(\varepsilon_t/\varepsilon_{t-1}, X_t\right) = h_t = c + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}. \end{array} \right. \quad (3.3)$$

3.1.4 Modèles ARCH-M (Engle-Lillien-Robbins [1987])

La formulation précédente n'est cependant pas bien adaptée aux problèmes financiers. En effet il existe des actifs plus ou moins risqués, et le rendement espéré de chaque actif doit refléter l'existence de ce risque : plus l'actif est risqué, plus le rendement anticipé devrait être élevé. Ceci conduit naturellement à des formulations du type :

$$\left\{ \begin{array}{l} \Phi(L)Y_t = A(L)X_t + \delta(L)h_t + \Theta(L)\varepsilon_t, \\ \text{avec les mêmes conditions qu'auparavant sur le processus } (\varepsilon_t). \end{array} \right.$$

Ce modèle diffère du précédent par la présence de la prime de risque $\delta(L)h_t$, qui dépend de variables prédéterminées *non observables*. Il est dit ARCH.M, c'est-à-dire « ARCH in mean » pour présence de la variance conditionnelle dans la moyenne conditionnelle.

Ce type de formulation plus proche de la théorie financière présente aussi d'autres avantages plus descriptifs. Il signifie en effet que la volatilité peut servir d'indicateur avancé pour évaluer le niveau moyen, c'est-à-dire l'évolution tendancielle. Or cette possibilité paraît compatible avec certaines données existantes.

3.1.5 Modèles autorégressifs à coefficients aléatoires (Bera-Lee [1989])

Certains modèles hétéroscélastiques peuvent être interprétés en terme de modèles autorégressifs à coefficients aléatoires. Cette interprétation a évidem-

ment un intérêt propre : elle va aussi permettre de mettre en évidence la forme peut être trop restrictive retenue pour la dynamique des moments d'ordre deux.

Considérons un processus autorégressif d'ordre p :

$$Y_t = \sum_{i=1}^p \varphi_{it} Y_{t-i} + u_t,$$

où (u_t) est un bruit blanc indépendant, de variance σ^2 .

Les coefficients sont susceptibles de varier dans le temps. Dans une version simple, ces coefficients peuvent être supposés temporellement indépendants, indépendants du processus d'erreur (u_t) et de même loi. Nous notons $E \varphi_{it} = \mu_i$, $V \varphi_t = \Sigma$. Le modèle précédent peut être intégré par rapport aux coefficients de façon à expliciter la loi des seules variables observables (Y_t) . Nous obtenons :

$$\left\{ \begin{array}{l} E(Y_t/Y_{t-1}) = \sum_{i=1}^p \mu_i Y_{t-i}, \\ V(Y_t/Y_{t-1}) = \sigma^2 + (Y_{t-1}, \dots, Y_{t-p}) \Sigma \begin{pmatrix} Y_{t-1} \\ \vdots \\ Y_{t-p} \end{pmatrix}. \end{array} \right. \quad (3.4)$$

Il s'agit d'un modèle autorégressif avec erreurs conditionnellement hétéroscédastiques. La formulation de la variance conditionnelle diffère cependant des formulations introduites précédemment.

- (i) Elle dépend des valeurs retardées du processus lui-même et non des valeurs retardées de l'innovation $\varepsilon_t = Y_t - E(Y_t/Y_{t-1})$, dès que les moyennes μ_i sont non nulles. Savoir si la volatilité doit être exprimée en fonction des niveaux de rendements ou des écarts de ceux-ci à une évolution moyenne est une question de fond. Elle conduit dans les applications financières à distinguer diverses notions de risque selon la définition retenue de l'évolution moyenne (voir par exemple les notions de risques systématique et spécifique).
- (ii) De plus nous voyons que la forme fonctionnelle de la variance conditionnelle, si elle est bien de type quadratique, comporte aussi des termes croisés du type $Y_{t-i}Y_{t-j}$, $i \neq j$, faisant intervenir des décalages différents.

3.1.6 Formulations non paramétriques

L'interprétation précédente montre que les formes fonctionnelles retenues pour la volatilité dans les modèles GARCH sont peut être trop restrictives. Dans le même ordre d'idée, on peut aussi remettre en question la formulation quadratique qui implique des réactions de la volatilité aux innovations passées indépendantes

du signe de ces innovations. Or on sait en pratique que les réactions en période de baisse des prix ou de hausse des prix peuvent être assez différentes.

Une approche naturelle consiste à retenir des formes peu contraintes pour les moyennes et variances conditionnelles. En pratique il faut essentiellement préciser les variables, dont les deux moments sont susceptibles de dépendre.

Ainsi, si nous supposons le processus markovien d'ordre p , nous pouvons écrire :

$$\begin{cases} E(Y_t/Y_{t-1}) = G(Y_{t-1}, \dots, Y_{t-p}), \\ V(Y_t/Y_{t-1}) = H(Y_{t-1}, \dots, Y_{t-p}). \end{cases} \quad (3.5)$$

Des modèles plus complexes peuvent être considérés, soit pour introduire des effets moyennes mobiles, soit pour faire dépendre la volatilité des valeurs retardées des innovations, soit pour introduire la volatilité dans la moyenne conditionnelle. Se restreignant à l'ordre un, une formulation prenant en compte ces aspects serait :

$$\begin{cases} E(Y_t/Y_{t-1}) = G(Y_{t-1}, \varepsilon_{t-1}, h_t), \\ \text{avec : } \varepsilon_t = Y_t - E(Y_t/Y_{t-1}), h_t = V(Y_t/Y_{t-1}), \\ V(Y_t/Y_{t-1}) = h_t = H(\varepsilon_{t-1}^2, h_{t-1}). \end{cases} \quad (3.6)$$

3.2 STATIONNARITÉ

L'étude des conditions de stationnarité faible ou forte est habituellement décrite pour des modèles GARCH. Certains résultats et principes de démonstrations peuvent cependant être facilement étendus à d'autres modèles conditionnellement hétéroscédastiques, en particulier aux modèles ARCH à seuils (voir Zakoïan [1991]).

3.2.1 Stationnarité faible

Considérons le modèle GARCH :

$$\begin{cases} Y_t = \varepsilon_t, \\ \text{avec } E(\varepsilon_t/\varepsilon_{t-1}) = 0 \\ E(\varepsilon_t^2/\varepsilon_{t-1}) = h_t = c + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}, \alpha_q \neq 0, \beta_p \neq 0. \end{cases}$$

Nous savons que la dynamique des variances peut s'écrire sous une forme ARMA :

$$\varepsilon_t^2 = c + \sum_{i=1}^{\text{Max}(p,q)} (\alpha_i + \beta_i) \varepsilon_{t-i}^2 + u_t - \sum_{j=1}^p \beta_j u_{t-j},$$

avec $E(u_t / \varepsilon_{t-1}) = 0$.

Prenant l'espérance de chacun des deux membres, nous en déduisons que :

$$E \varepsilon_t^2 = c + \sum_{i=1}^{\text{Max}(p,q)} (\alpha_i + \beta_i) E \varepsilon_{t-i}^2.$$

La suite des variances $(E \varepsilon_t^2) = (\sigma_t^2)$ satisfait donc une équation de récurrence linéaire. Le processus (ε_t) , qui est déjà une différence de martingale, est asymptotiquement stationnaire faible dès que les racines du polynôme caractéristique :

$$1 - \sum_{i=1}^{\text{Max}(p,q)} (\alpha_i + \beta_i) z^i,$$

sont à l'extérieur du disque unité.

Propriété 3.1

Etant donné un processus GARCH avec des coefficients α_i, β_i positifs, la condition

$$1 - \sum_{i=1}^{\text{Max}(p,q)} (\alpha_i + \beta_i) z^i = 0 \implies |z| > 1,$$

est équivalente à :

$$\sum_{i=1}^{\text{Max}(p,q)} (\alpha_i + \beta_i) < 1.$$

Preuve

Notons :

$$\alpha(z) = \sum_{i=1}^q \alpha_i z^i, \beta(z) = \sum_{i=1}^p \beta_i z^i.$$

Condition nécessaire : Supposons $\alpha(1) + \beta(1) \geq 1$. La fonction $1 - \alpha(z) - \beta(z)$ est continue, strictement positive pour $z = 0$, négative pour $z = 1$. Elle admet donc un zéro à l'intérieur du disque unité.

Condition suffisante : Inversement, supposons $\alpha(1) + \beta(1) \leq 1$. Si z_0 de module inférieur à 1 était racine du polynôme caractéristique, on aurait :

$$\begin{aligned}
 1 &= \sum_{i=1}^{\text{Max}(p,q)} (\alpha_i + \beta_i) z_0^i \\
 &= \left| \sum_{i=1}^{\text{Max}(p,q)} (\alpha_i + \beta_i) z_0^i \right| \\
 &\leq \sum_{i=1}^{\text{Max}(p,q)} |\alpha_i + \beta_i| |z_0|^i \\
 &\leq \sum_{i=1}^{\text{Max}(p,q)} |\alpha_i + \beta_i| \\
 &= \sum_{i=1}^{\text{Max}(p,q)} (\alpha_i + \beta_i) \\
 &< 1, \text{ d'où une contradiction.}
 \end{aligned}$$

□

Les conditions de positivité des coefficients c , α_i , β_i assurent évidemment la positivité de la variance conditionnelle h_t . Cette dernière peut malgré tout être positive p.s. dans des cas où certains coefficients α_i et β_i seraient (faiblement) négatifs. On a l'habitude de négliger ces cas et d'assimiler condition de stationnarité et condition sur la somme des coefficients.

3.2.2 Valeurs rencontrées en pratique

A titre d'illustration, nous présentons ci-dessous les résultats d'estimation relatifs à des indices de marché (Grar [1992]). Quatre indices ont été considérés à partir de la base de données de la Société des Bourses Françaises. Cette base permet de calculer les rentabilités quotidiennes de 1914 valeurs françaises côtées sur la bourse de Paris ou sur les bourses de Province, et a été utilisée pour la période du 3 Janvier 1977 au 29 Décembre 1990.

Les portefeuilles agrégés se différencient par le type de pondération (équipondération ou pondération par capitalisation boursière) et par le traitement des dividendes (avec ou sans réinvestissement des dividendes). Ceci conduit à quatre portefeuilles proxy de l'indice de marché, dans l'ordre :

- (I) portefeuille équipondéré avec réinvestissement des dividendes ;
- (II) portefeuille équipondéré sans réinvestissement des dividendes ;

- (III) portefeuille pondéré avec réinvestissement des dividendes ;
 (IV) portefeuille pondéré sans réinvestissement des dividendes.

Nous donnons dans le tableau 3.1 quelques résumés de la distribution historique des rentabilités de ces portefeuilles.

Tableau 3.1: Distribution empirique des rentabilités

	(I)	(II)	(III)	(IV)
	Avec divid.	Sans divid.	Avec divid.	Sans divid.
Moyenne	0,54 10^{-3}	0,48 10^{-3}	0,41 10^{-3}	0,34 10^{-3}
Variance	0,23 10^{-4}	0,85 10^{-4}	0,23 10^{-4}	0,86 10^{-4}
Coefficient d'asymétrie	-1,60	-1,00	-1,58	-0,99
Kurtosis	12,06	9,76	11,88	9,66

On remarque immédiatement que ces chiffres sont incompatibles avec la normalité de cette distribution historique à la fois du fait de l'asymétrie négative et des fortes queues de distribution, la kurtosis prenant des valeurs significativement supérieures à 3 (valeur correspondant à la loi normale) (Ce problème sera discuté dans la section 3.3). Par ailleurs bien que la moyenne des rentabilités soit proche de zéro, on voit qu'elle est toujours plus grande lorsque les dividendes sont pris en compte. Prendre en compte les dividendes revient à augmenter au fur et à mesure les pondérations des actifs versant des dividendes importants, ce qui explique évidemment cet effet. Accepter l'hypothèse que la moyenne des rentabilités est nulle, risque donc de dépendre beaucoup de la façon de calculer les rentabilités.

Finalement prendre en compte les dividendes a tendance du fait des changements de pondérations à diminuer l'hétérogénéité, donc à diminuer la variance, ce qui est constaté sur le moment d'ordre deux.

Pour chacun de ces portefeuilles, un modèle de type GARCH(1,1) avec effet de translation a été ajusté sur les rentabilités. Ce modèle est :

$$\begin{cases} Y_t = a + \varepsilon_t, \\ E(\varepsilon_t / \varepsilon_{t-1}) = 0, \\ V(\varepsilon_t / \varepsilon_{t-1}) = h_t = c + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}. \end{cases}$$

Il a été estimé par pseudo-maximum de vraisemblance (voir chapitre 4).

Tableau 3.2: Modèles estimés

Terme	(I)	(II)	(III)	(IV)
	Avec divid.	Sans divid.	Avec divid.	Sans divid.
a	0,36 10 ⁻³ (7, 6)	-0,53 10 ⁻³ (5, 2)	0,27 10 ⁻³ (5, 4)	-0,72 10 ⁻³ (11, 6)
c	0,61 10 ⁻⁶ (23, 6)	0,42 10 ⁻⁵ (23, 1)	0,75 10 ⁻⁶ (20, 5)	0,80 10 ⁻⁶ (8, 4)
α	0,22 (20, 5)	0,21 (14, 4)	0,28 (19, 4)	0,21 (27, 3)
β	0,72 (45, 4)	0,73 (44, 9)	0,71 (43, 1)	0,73 (45, 8)

Les valeurs du terme constant qui s'interprètent comme les moyennes conditionnelles des rentabilités sont significativement différentes de zéro, ce qui conduit à rejeter l'hypothèse de marche aléatoire pour la suite des valeurs des portefeuilles associées à ces rentabilités. On observe par ailleurs le changement de signe de ce coefficient estimé selon que les dividendes sont ou non pris en compte.

Surtout on note une hétéroscédasticité conditionnelle significative. Celle-ci est telle que la somme des deux coefficients α et β, égale respectivement à 1, 0,94, 0,99 et 0,94, est proche de 1, donc de la condition limite pour avoir la stationnarité faible. Il apparaît donc très important d'étudier plus en détail ce cas.

Finalement nous donnons, à titre d'illustration (voir les deux pages suivantes), l'évolution de la variance conditionnelle estimée \hat{h}_t pour les quatre portefeuilles précédents.

3.2.3 Stationnarité forte

Des racines unitaires sont ainsi souvent mises en évidence dans la dynamique des moments conditionnels d'ordre deux et il est nécessaire d'étudier ce cas limite, où il ne devrait plus y avoir stationnarité faible du processus. L'intuition et l'analogie avec le cas de racines unitaires dans les modèles ARMA conduisent à penser que le processus devrait alors présenter des non-stationnarités. En fait il n'en est rien : la stationnarité sera conservée, mais les valeurs trop fortes des coefficients autorégressifs conduiront à la non-existence des moments d'ordre deux. Cette propriété a été initialement montrée par Nelson [1990] pour un modèle GARCH(1,1) et nous suivrons sa démarche.

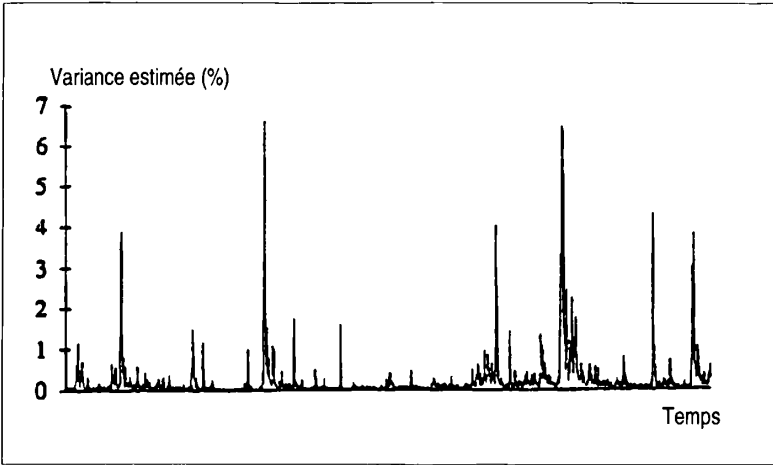


Figure 3.1

Évolution de la volatilité: Indice équi pondéré avec dividende (GARCH(1,1))

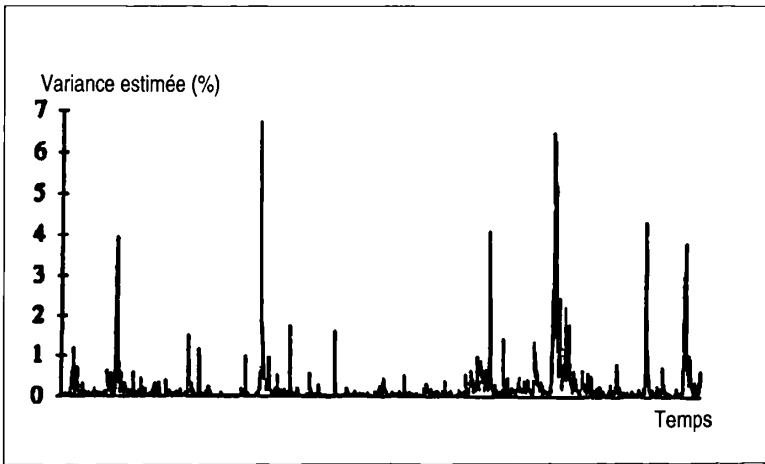


Figure 3.2

Évolution de la volatilité: Indice équi pondéré sans dividende (GARCH(1,1))

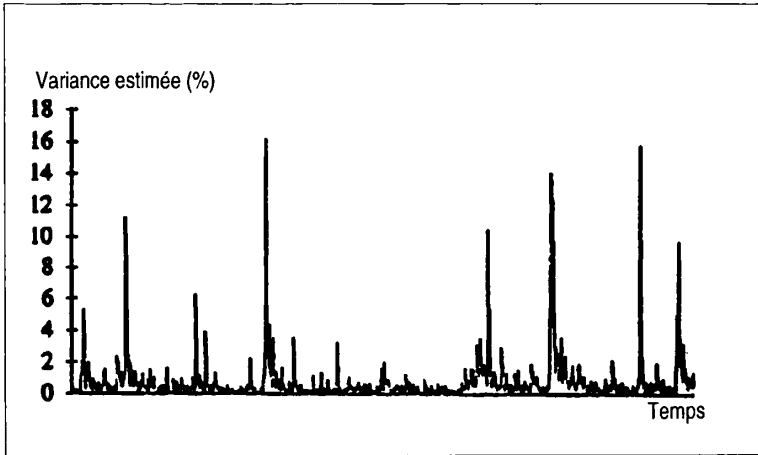


Figure 3.3

Évolution de la volatilité : Indice pondéré avec dividende (GARCH(1,1))

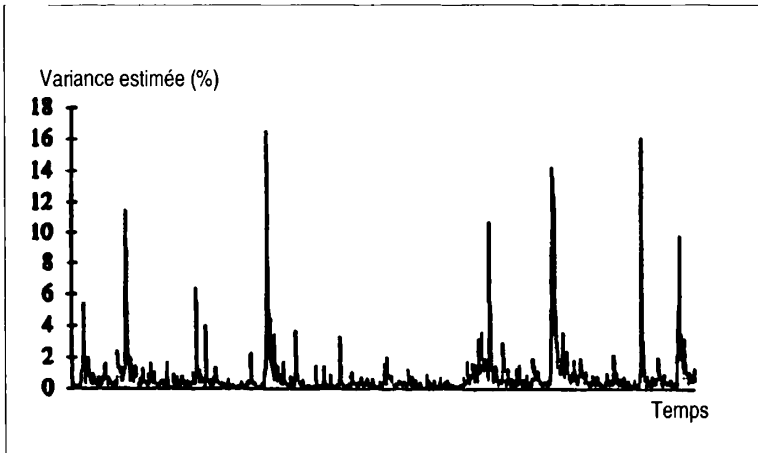


Figure 3.4

Évolution de la volatilité : Indice pondéré sans dividende (GARCH(1,1))

Considérons donc un processus satisfaisant :

$$\begin{cases} Y_t = \varepsilon_t, \\ \varepsilon_t / \varepsilon_{t-1} \sim N(0, h_t), \\ \text{avec : } h_t = c + \beta h_{t-1} + \alpha \varepsilon_{t-1}^2, \quad \alpha \geq 0, \beta \geq 0, c \geq 0, \end{cases}$$

où les paramètres α, β sont susceptibles de prendre toutes les valeurs, y compris des valeurs pour lesquelles $\alpha + \beta \geq 1$.

La condition sur la volatilité peut être réécrite en introduisant le bruit blanc gaussien réduit $Z_t = h_t^{-1/2} \varepsilon_t$.

Nous avons :

$$h_t = c + (\beta + \alpha Z_{t-1}^2) h_{t-1},$$

ou par remplacement récursif à partir d'une date initiale zéro :

$$h_t = h_0 \prod_{i=1}^t (\beta + \alpha Z_{t-i}^2) + c \left[1 + \sum_{k=1}^{t-1} \prod_{i=1}^k (\beta + \alpha Z_{t-i}^2) \right].$$

L'étude du processus initial :

$$Y_t = \varepsilon_t = h_t^{1/2} Z_t,$$

se ramène alors intuitivement à celle du processus h_t .

Nous allons nous concentrer sur l'effet de la valeur initiale h_0 , c'est-à-dire sur le terme $h_0 \prod_{i=1}^t (\beta + \alpha Z_{t-i}^2)$ figurant dans h_t . Ceci nous fournira une idée du résultat et nous renvoyons à Nelson [1990] pour une preuve rigoureuse de la propriété.

i) Persistance en moyenne

On peut d'abord s'intéresser à l'effet multiplicateur moyen de h_0 sur h_t , c'est-à-dire à :

$$\begin{aligned} E \left(\prod_{i=1}^t (\beta + \alpha Z_{t-i}^2) \right) &= \prod_{i=1}^t E (\beta + \alpha Z_{t-i}^2) \\ &= (\beta + \alpha)^t. \end{aligned}$$

Nous constatons que cet effet disparaît asymptotiquement si $\alpha + \beta < 1$, qu'il persiste en moyenne si $\alpha + \beta = 1$ et qu'il est explosif si $\alpha + \beta > 1$.

De plus, si le processus est fortement stationnaire, nous avons :

$$E h_t > E h_0 (\alpha + \beta)^\tau, \forall \tau,$$

de sorte que si $\alpha + \beta > 1$ nous en déduisons que $E h_t = V Y_t = +\infty$, c'est-à-dire que le moment d'ordre deux n'existe pas.

ii) Persistance trajectoire par trajectoire

En revanche, si nous considérons l'effet multiplicateur lui-même, nous avons :

$$\begin{aligned} \prod_{i=1}^t (\beta + \alpha Z_{t-i}^2) &= \exp \sum_{i=1}^t \text{Log} [\beta + \alpha Z_{t-i}^2] \\ &= \exp \left\{ t \frac{1}{t} \sum_{i=1}^t \text{Log} (\beta + \alpha Z_{t-i}^2) \right\} \end{aligned}$$

Comme $\frac{1}{t} \sum_{i=1}^t \text{Log} (\beta + \alpha Z_{t-i}^2)$ tend vers $E \text{Log} (\beta + \alpha Z^2)$, nous voyons que l'effet multiplicateur tend p.s vers zéro si et seulement si $E \text{Log} (\beta + \alpha Z^2) < 0$. Dans le cas contraire il y a persistance trajectoire pour trajectoire.

iii) Comparaison des deux conditions

D'après l'inégalité de convexité :

$$E \text{Log} [\alpha + \beta Z^2] < \text{Log} E (\alpha + \beta Z^2) = \text{Log}(\alpha + \beta).$$

Ainsi la condition $\alpha + \beta < 1$ entraîne la condition $E \text{Log} [\alpha + \beta Z^2] < 0$. La non-persistance en moyenne entraîne la non-persistance trajectoire par trajectoire.

La variable Z suivant une loi normale centrée réduite, la valeur de $E \text{Log} [\alpha + \beta Z^2]$ est :

$$\begin{aligned} E \text{Log} (\alpha + \beta Z^2) &= \\ &= \text{Log} 2\alpha + \Psi(1/2) + [2\Pi\beta/\alpha]^{1/2} {}_1F_1(1/2; 3/2; \beta/2\alpha) \\ &\quad - \beta/\alpha {}_2F_2(1, 1; 2, 3/2; \beta/2\alpha), \end{aligned}$$

où Ψ est la fonction Ψ d'Euler, ${}_1F_1$ la fonction hypergéométrique, ${}_2F_2$ la fonction hypergéométrique généralisée.

Dans le plan (α, β) les formes des diverses régions sont celles représentées dans la figure 3.2.

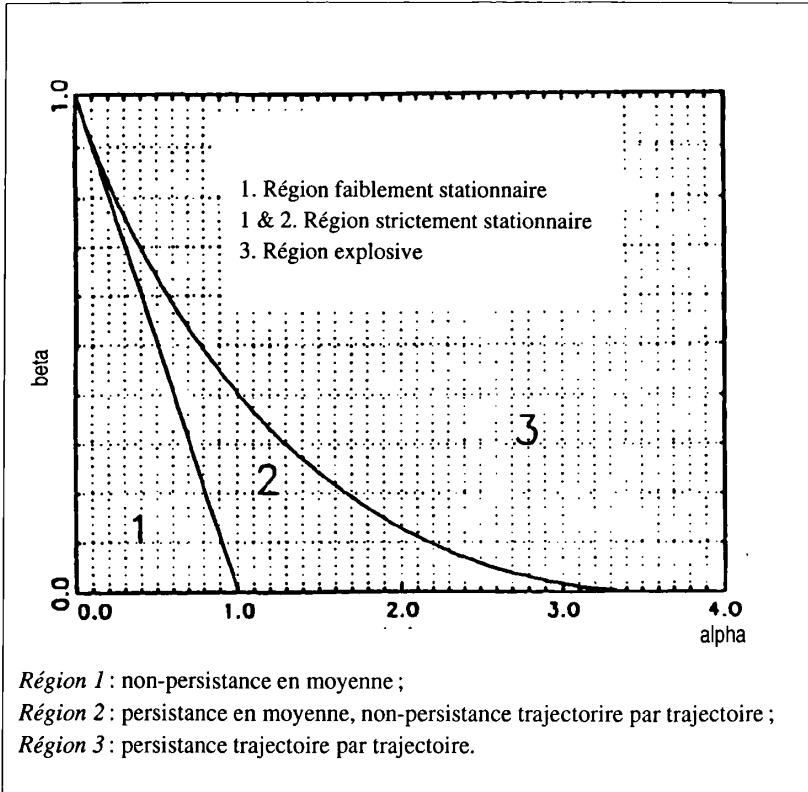


Figure 3.5: Régions de persistance et de stationnarité

iv) Critères de stationnarité

Ces régions ont aussi des interprétations en terme de stationnarité.

Propriété 3.2 : (Voir Nelson [1990])

Le processus (h_t^*, Y_t^*) avec

$$h_t^* = c \left[1 + \sum_{k=0}^{\infty} \prod_{i=1}^k (\beta + \alpha Z_{t-i}^2) \right],$$

$$Y_t^* = h_t^{*1/2} Z_t,$$

est fortement stationnaire si et seulement si :

$$E \text{Log} [\beta + \alpha Z_t^2] < 0.$$

La partition entre les trois régions s'interprète alors de la façon suivante :

région 1 : Y_t est fortement stationnaire et son moment d'ordre deux existe ;

région 2 : Y_t est fortement stationnaire et son moment d'ordre deux n'existe pas ;

région 3 : Y_t n'est pas fortement stationnaire et le moment d'ordre deux n'existe pas.

Les processus pour lesquels $\alpha + \beta = 1$ sont appelés **IGARCH** (I comme intégré). Les queues des lois marginales sont alors suffisamment épaisses pour empêcher l'existence de la variance.

3.3 QUEUES DE DISTRIBUTION

Nous avons vu que pour certaines valeurs des paramètres, les variances conditionnelles pouvaient être définies sans que la variance marginale existe. Ceci résulte de l'accroissement des queues de distribution par agrégation temporelle, lorsque la série est conditionnellement hétéroscédastique. Ce **phénomène** est dit **leptokurtique**, l'importance des queues étant souvent mesurée par l'intermédiaire de la kurtosis, rapport du moment centré d'ordre quatre par le carré de la variance. Plus la kurtosis est grande, plus les queues sont épaisses, ceci pour des distributions suffisamment régulières.

3.3.1 Lien entre kurtosis conditionnelle et kurtosis marginale

Considérons un processus $(Y_t) = (\varepsilon_t)$ conditionnellement centré et conditionnellement hétéroscédastique :

$$E(\varepsilon_t / \varepsilon_{t-1}) = 0$$

$$E(\varepsilon_t^2 / \varepsilon_{t-1}) = h_t.$$

La kurtosis conditionnelle est :

$$k_t = \frac{E(\varepsilon_t^4 / \varepsilon_{t-1})}{[E(\varepsilon_t^2 / \varepsilon_{t-1})]^2}.$$

Etudions alors le moment marginal d'ordre quatre. Nous avons :

$$\begin{aligned} E(\varepsilon_t^4) &= E\left[E(\varepsilon_t^4 / \varepsilon_{t-1})\right] \\ &= E[k_t h_t^2] \\ &= E k_t E h_t^2 + \text{Cov}(k_t, h_t^2) \end{aligned}$$

$$\begin{aligned} &\geq \mathbb{E} k_t (\mathbb{E} h_t)^2 + \text{Cov} (k_t, h_t^2) \\ &= \mathbb{E} k_t (\mathbb{E} \varepsilon_t^2)^2 + \text{Cov} (k_t, h_t^2). \end{aligned}$$

La kurtosis marginale $k = \frac{\mathbb{E}(\varepsilon_t^4)}{(\mathbb{E}\varepsilon_t^2)^2}$ est donc telle que :

$$k - \mathbb{E}(k_t) \geq \text{Cov} \left[k_t, \left(\frac{h_t}{\mathbb{E} h_t} \right)^2 \right]. \quad (3.7)$$

Selon les périodes, l'actif peut être plus risqué (h_t et k_t sont alors simultanément grands) ou moins risqué (h_t et k_t sont simultanément petits). On s'attend à ce que h_t et k_t soient en liaison plutôt croissante et donc à ce que $\text{Cov}(k_t, h_t^2) \geq 0$. La kurtosis marginale k est alors supérieure à la moyenne des kurtosis conditionnelles.

Remarque : Lorsque la loi conditionnelle est normale, la kurtosis conditionnelle k_t est une constante égale à 3 et l'inégalité se réduit à $k \geq 3$. La loi marginale a des queues plus épaisses que celles de la loi normale. De plus, on a :

$$\begin{aligned} k - 3 &= k - \mathbb{E} k_t \\ &= \mathbb{E} k_t [\mathbb{E} h_t^2 - (\mathbb{E} h_t)^2] \\ &= 3 \mathbb{V} h_t. \end{aligned}$$

L'écart entre kurtosis marginale et moyenne des kurtosis conditionnelles est directement lié à la variabilité des volatilités h_t . En particulier $k = 3$ si et seulement si $\mathbb{V} h_t = 0$, c'est-à-dire si et seulement si la variance conditionnelle est constante (cas conditionnellement homoscédastique).

3.3.2 Kurtosis et choix de l'unité de temps

Les résultats précédents peuvent être généralisés de la façon suivante. Considérons un processus $(Y_t) = (\varepsilon_t)$ conditionnellement centré et hétéroscédastique. Si nous retenons une unité de temps H à valeurs entières, nous pouvons introduire le processus $\tilde{Y}^{(H)} = Y_{t-H}$, $t \in \mathbb{Z}$, consistant à effectuer les observations de H en H . Pour chaque unité de temps nous pouvons considérer la loi conditionnelle de $\tilde{Y}_t^{(H)}$ sachant $\tilde{Y}_{t-1}^{(H)}$ et la kurtosis conditionnelle correspondante $k_t^{(H)}$. L'étude du paragraphe précédent a simplement consisté à comparer les kurtosis avec unité $H = 1$ et unité $H = +\infty$. On peut cependant étudier de façon plus précise l'évolution de $k_t^{(H)}$ avec l'unité H . Nous menons cette étude pour un processus ARCH (1) défini par :

$$Y_t = \sqrt{a_0 + a_1 Y_{t-1}^2} \varepsilon_t,$$

où (ε_t) est un bruit blanc gaussien centré réduit. Ce processus étant markovien d'ordre un, il en est de même des divers processus $(\tilde{Y}_t^{(H)})$. Les moments conditionnels du processus initial sont tels que :

$$\begin{aligned} E\left(Y_t^2/Y_{t-1}\right) &= a_0 + a_1 Y_{t-1}^2, \\ E\left(Y_t^4/Y_{t-1}\right) &= (a_0 + a_1 Y_{t-1}^2)^2 E\varepsilon_t^4 \\ &= 3a_0^2 + 6a_0 a_1 Y_{t-1}^2 + 3a_1^2 Y_{t-1}^4. \end{aligned}$$

Sous forme matricielle ce système s'écrit :

$$\begin{bmatrix} E\left(Y_t^2 - m_2/Y_{t-1}\right) \\ E\left(Y_t^4 - m_4/Y_{t-1}\right) \end{bmatrix} = \begin{bmatrix} a_1 & 0 \\ 6a_0 a_1 & 3a_1^2 \end{bmatrix} \begin{bmatrix} Y_{t-1}^2 - m_2 \\ Y_{t-1}^4 - m_4 \end{bmatrix},$$

où

$$m_2 = \frac{a_0}{1 - a_1} \text{ et } m_4 = \frac{3a_0^2}{(1 - a_1)^2} \frac{1 - a_1^2}{1 - 3a_1^2}$$

sont les moments marginaux d'ordre deux et quatre.

Par itération nous en déduisons :

$$\begin{aligned} \begin{pmatrix} E\left[Y_t^2 - m_2/Y_{t-H}\right] \\ E\left[Y_t^4 - m_4/Y_{t-H}\right] \end{pmatrix} &= \begin{bmatrix} a_1 & 0 \\ 6a_0 a_1 & 3a_1^2 \end{bmatrix}^H \begin{bmatrix} Y_{t-H}^2 - m_2 \\ Y_{t-H}^4 - m_4 \end{bmatrix} \\ &= \begin{bmatrix} a_1^H & 0 \\ C_H & 3^H a_1^{2H} \end{bmatrix} \begin{bmatrix} Y_{t-H}^2 - m_2 \\ Y_{t-H}^4 - m_4 \end{bmatrix}, \end{aligned}$$

avec :

$$C_H = \frac{6a_0}{3a_1 - 1} \left\{ (3a_1 - 1 - 18a_0 a_1^2) a_1^H + 6a_0 a_1 3^H a_1^{2H} \right\}.$$

Les deux moments conditionnels s'écrivent donc :

$$\begin{cases} E\left[Y_t^2/Y_{t-H}\right] = (1 - a_1^H) m_2 + a_1^H Y_{t-H}^2 \\ E\left[Y_t^4/Y_{t-H}\right] = (1 - 3^H a_1^{2H}) m_4 - C_H m_2 + C_H Y_{t-H}^2 + 3^H a_1^{2H} Y_{t-H}^4 \end{cases},$$

La kurtosis avec unité de temps H :

$$k_t^{(H)} = \frac{E\left(Y_{tH}^4/Y_{(t-1)H}\right)}{\left[E\left(Y_{tH}^2/Y_{(t-1)H}\right)\right]^2},$$

est pour H quelconque généralement non constante (alors qu'elle l'était dans le paragraphe précédent pour les cas limites $H = 1$ et $H = +\infty$). Il est donc

possible que, pour certaines valeurs de H et certaines trajectoires passées, la kurtosis augmente.

3.3.3 Normalité conditionnelle et choix de l'unité de temps

La valeur de la kurtosis n'étant pas invariante par changement de l'unité de temps, il en est de même de la condition de normalité conditionnelle. S'il y a une normalité conditionnelle pour des valeurs observées de Y_t , cette normalité conditionnelle n'est plus satisfaite pour des valeurs observées de H en $H \neq 1$ dès qu'il y a hétéroscédasticité conditionnelle.

Ceci conduit à se poser la question suivante : supposons que nous disposions d'observations (\tilde{Y}_t) pour lesquelles la kurtosis est nettement supérieure à 3, existe-t-il une unité de temps sous-jacente, plus petite que un et pour laquelle on pourrait faire l'hypothèse de normalité conditionnelle ? On ne peut évidemment répondre à une telle question que dans le cadre d'un modèle plus précis, par exemple en supposant que :

$$\tilde{Y}_t = \tilde{Y}_t^{(H)} = Y_{tH},$$

où le processus (Y_t) sous-jacent admettrait une formulation ARCH(1), conditionnellement gaussienne. Comment pourrait-on alors déterminer la valeur H ? Une démarche pragmatique serait basée sur les formes des moments conditionnels et les régressions suivantes :

- a) Effectuer la régression de \tilde{Y}_t^2 sur 1, \tilde{Y}_{t-1}^2 , ce qui fournit des coefficients estimés :

$$\tilde{Y}_t^2 \simeq \hat{\alpha}_1 + \hat{\alpha}_2 \tilde{Y}_{t-1}^2.$$

- b) Effectuer la régression de \tilde{Y}_t^4 sur 1, \tilde{Y}_{t-1}^2 , \tilde{Y}_{t-1}^4 , ce qui fournit des coefficients estimés :

$$\tilde{Y}_t^4 \simeq \hat{\beta}_1 + \hat{\beta}_2 \tilde{Y}_{t-1}^2 + \hat{\beta}_3 \tilde{Y}_{t-1}^4.$$

- c) Si le modèle supposé est valable, on devrait alors d'après (3.8) avoir :

$$\hat{\alpha}_2 \simeq a_1^H, \hat{\beta}_3 \simeq 3^H a_1^{2H},$$

où

$$H \simeq \frac{\text{Log } \hat{\beta}_3 - 2 \text{ log } \hat{\alpha}_2}{\text{Log } 3}.$$

Cette démarche est essentiellement descriptive. Cependant elle a l'avantage de montrer que pour l'étude de modèles conditionnellement hétéroscédastiques ne considérer que la dynamique des moments d'ordre deux est vraisemblablement trop restrictif. Une analyse même simple de la dynamique des moments d'ordre supérieur, ici quatre, peut aussi se révéler informative.

CHAPITRE 4

ESTIMATION, PRÉVISION ET TESTS

Modèles univariés

par

Christian GOURIEROUX

Les modèles introduits dans le chapitre précédent reposent sur des formulations des moyennes et variances conditionnelles.

En pratique celles-ci sont souvent paramétrées de façon que la moyenne conditionnelle $m_t(\theta)$ et la variance conditionnelle $h_t(\theta)$ apparaissent comme des fonctions de paramètres inconnus et de valeurs passées du processus. La connaissance de ces moments ne suffit cependant pas sans hypothèse supplémentaire à caractériser la loi conditionnelle du processus. Les méthodes d'estimation semi-paramétriques utilisées en pratique sont essentiellement de trois types :

- méthode du pseudo-maximum de vraisemblance, reposant sur une fonction critère déduite de la loi normale,
- méthode en deux étapes utilisant certaines spécificités des formulations retenues pour les moments conditionnels,
- méthode par variables instrumentales.

Diverses procédures de test peuvent alors être associées à ces méthodes d'estimation et serviront notamment pour étudier les hypothèses d'homoscédasticité conditionnelle ou de marche aléatoire.

La fin du chapitre est consacrée à une présentation de démarches de type non paramétrique.

4.1 ESTIMATION LORSQUE LES MOMENTS SONT PARAMÉTRÉS

Dans tout ce chapitre nous indexons par t les espérances et variances calculées par rapport à la vraie loi du processus, et supposons celui-ci fortement stationnaire. Les spécifications des moyennes et variances conditionnelles sont :

$$\begin{cases} \mathbb{E}_0 \left(Y_t / \underline{Y}_{t-1} \right) = m_t(\theta_0), \\ \mathbb{V}_0 \left(Y_t / \underline{Y}_{t-1} \right) = h_t(\theta_0). \end{cases}$$

où θ_0 est la vraie valeur inconnue du paramètre θ appartenant à Θ inclus dans \mathbb{R}^k . Nous serons amenés à introduire les processus conditionnellement centrés, réduits :

$$u_t(\theta) = \frac{Y_t - m_t(\theta)}{h_t(\theta)^{1/2}},$$

et en particulier celui correspondant à la vraie valeur θ_0 du paramètre :

$$u_t = u_t(\theta_0).$$

Ce processus est tel que :

$$\begin{cases} \mathbb{E}_0 \left(u_t / \underline{Y}_{t-1} \right) = 0, \\ \mathbb{V}_0 \left(u_t / \underline{Y}_{t-1} \right) = 1. \end{cases}$$

Nous notons :

$$\begin{cases} M_{3t}(\theta) = \mathbb{E}_0 \left(u_t(\theta)^3 / \underline{Y}_{t-1} \right), \\ K_t(\theta) = \mathbb{E}_0 \left(u_t(\theta)^4 / \underline{Y}_{t-1} \right), \end{cases}$$

les moments d'ordre supérieur.

4.1.1 Méthodes du pseudo-maximum de vraisemblance

Dans cette approche, on recherche un estimateur de θ par une méthode du maximum de vraisemblance appliquée comme si le processus (u_t) était un bruit blanc gaussien réduit. On montre alors que cette erreur de spécification éventuelle sur la loi du processus (u_t) n'a pas de conséquence sur les propriétés de convergence de l'estimateur, mais qu'il est seulement nécessaire d'effectuer une correction au niveau de l'évaluation de la précision asymptotique. De façon à ne pas compliquer la présentation et parce que les résultats ne sont pas spécifiques des modèles ARCH, nous n'insistons ni sur les démonstrations des propriétés asymptotiques, ni sur les conditions de régularité sous lesquelles elles sont valides (voir White [1982], Gouriéroux, Monfort et Trognon [1984]).

Ces conditions de régularité sont toujours de trois types :

1. des conditions de stationnarité forte et de mélangeance du processus (Y_t) ,
2. des conditions d'existence des dérivées et des moments apparaissant dans les diverses formules,
3. des conditions d'identifiabilité du paramètre θ , qui doit pouvoir être retrouvé sans ambiguïté à partir des deux premiers moments conditionnels.

Définition 4.1

L'estimateur du pseudo-maximum de vraisemblance $\hat{\theta}_T$ du paramètre θ est une solution du problème :

$$\text{Max}_{\theta \in \Theta} \text{Log } L_T(\theta)$$

avec :

$$\text{Log } L_T(\theta) = -\frac{T}{2} \text{Log } 2\pi - \frac{1}{2} \sum_{t=1}^T \text{Log } h_t(\theta) - \frac{1}{2} \sum_{t=1}^T \frac{(Y_t - m_t(\theta))^2}{h_t(\theta)}$$

Cet estimateur est en pratique recherché en résolvant les conditions du premier ordre :

$$\frac{\partial \text{Log } L_T(\hat{\theta}_T)}{\partial \theta} = 0.$$

Nous avons :

$$\begin{aligned} \frac{\partial \text{Log } L_T(\theta)}{\partial \theta} &= -\frac{1}{2} \sum_{t=1}^T \frac{1}{h_t(\theta)} \frac{\partial h_t(\theta)}{\partial \theta} + \frac{1}{2} \sum_{t=1}^T \frac{[Y_t - m_t(\theta)]^2}{h_t(\theta)^2} \frac{\partial h_t(\theta)}{\partial \theta} \\ &\quad + \sum_{t=1}^T \frac{Y_t - m_t(\theta)}{h_t(\theta)} \frac{\partial m_t(\theta)}{\partial \theta}. \end{aligned} \tag{4.1}$$

Introduisant les résidus centrés, réduits :

$$\hat{u}_t = u_t(\hat{\theta}_T) = \frac{Y_t - m_t(\hat{\theta}_T)}{h_t(\hat{\theta}_T)^{1/2}}, \tag{4.2}$$

la condition du premier ordre s'écrit :

$$\frac{1}{2} \sum_{t=1}^T \frac{1}{h_t(\hat{\theta}_T)} \frac{\partial}{\partial \theta} h_t(\hat{\theta}_T) (\hat{u}_t^2 - 1) + \sum_{t=1}^T \frac{1}{h_t(\hat{\theta}_T)^{1/2}} \frac{\partial m_t}{\partial \theta}(\hat{\theta}_T) \hat{u}_t = 0 \tag{4.3}$$

Elle apparaît comme une condition d'orthogonalité entre les résidus d'ordre un \hat{u}_t , et d'ordre deux $\hat{u}_t^2 - 1$ avec certains régresseurs (estimés)

$$\frac{1}{h_t(\hat{\theta}_T)^{1/2}} \frac{\partial m_t}{\partial \theta}(\hat{\theta}_T) \quad \text{et} \quad \frac{1}{h_t(\hat{\theta}_T)} \frac{\partial}{\partial \theta} h_t(\hat{\theta}_T)$$

respectivement.

Propriété 4.1

Sous les conditions de régularité :

1. Il existe une solution $\hat{\theta}_T$ qui converge vers la vraie valeur θ_0 .
2. L'estimateur est asymptotiquement normal :

$$\sqrt{T} (\hat{\theta}_T - \theta_0) \xrightarrow{\text{loi}} N [0, J^{-1} I J^{-1}],$$

avec

$$J = E_0 \left[\frac{\partial^2 \text{Log } \ell_t(Y; \theta_0)}{\partial \theta \partial \theta'} \right],$$

$$I = E_0 \left[\frac{\partial \text{Log } \ell_t(Y; \theta_0)}{\partial \theta} \frac{\partial \text{Log } \ell_t(Y; \theta_0)}{\partial \theta'} \right],$$

où ℓ_t est le t^{e} terme de la vraisemblance.

La méthode du pseudo-maximum de vraisemblance reposant sur une distribution, la distribution normale, éventuellement mal spécifiée, les deux expressions usuelles I et J de la matrice d'information ne sont en général pas égales. Elles peuvent être explicitées dans notre cas particulier. Un calcul simple donne :

$$J = E_0 \left\{ \frac{1}{h_t(\theta_0)} \frac{\partial m_t(\theta_0)}{\partial \theta} \frac{\partial m_t(\theta_0)}{\partial \theta'} + \frac{1}{2h_t(\theta_0)^2} \frac{\partial h_t(\theta_0)}{\partial \theta} \frac{\partial h_t(\theta_0)}{\partial \theta'} \right\},$$

$$I = E_0 \left\{ \frac{1}{4} \frac{1}{h_t(\theta_0)^2} \frac{\partial h_t(\theta_0)}{\partial \theta} \frac{\partial h_t(\theta_0)}{\partial \theta'} [K_t(\theta_0) - 1] \right. \\ \left. + \frac{1}{h_t(\theta_0)} \frac{\partial m_t(\theta_0)}{\partial \theta} \frac{\partial m_t(\theta_0)}{\partial \theta'} \right. \\ \left. + \frac{1}{2h_t(\theta_0)^{3/2}} \left[\frac{\partial h_t(\theta_0)}{\partial \theta} \frac{\partial m_t(\theta_0)}{\partial \theta'} + \frac{\partial m_t(\theta_0)}{\partial \theta} \frac{\partial h_t(\theta_0)}{\partial \theta'} \right] M_{3t}(\theta_0) \right\}.$$

Ces matrices sont en pratique approchées en remplaçant les espérances par les moyennes empiriques sur observations et la vraie valeur θ_0 par son estimateur $\hat{\theta}_T$.

4.1.2 Exemples

Les formules donnant les précisions asymptotiques des estimateurs du pseudo-maximum de vraisemblance peuvent se simplifier pour certains modèles particuliers.

i) Modèles conditionnellement normaux

Si les variables u_t suivent effectivement des lois normales, la méthode coïncide avec la maximum de vraisemblance. On a $K_t(\theta_0) = 3$, $M_{3t}(\theta_0) = 0$, et on vérifie directement que $I = J$ et

$$V_{as} \left[\sqrt{T}(\hat{\theta}_T - \theta_0) \right] = I^{-1} = J^{-1}.$$

ii) Paramétrages indépendants de la moyenne et de la variance

Un autre cas simple est celui où le paramètre θ peut se décomposer en

$$\theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix},$$

α n'apparaissant que dans la moyenne et β que dans la variance :

$$\begin{aligned} m_t(\theta) &= m_t(\alpha), \\ h_t(\theta) &= h_t(\beta) \text{ (avec un abus de notation).} \end{aligned}$$

Nous avons alors :

$$J = \begin{bmatrix} E_0 \left[\frac{1}{h_t(\beta_0)} \frac{\partial m_t(\alpha_0)}{\partial \alpha} \frac{\partial m_t(\alpha_0)}{\partial \alpha'} \right] & 0 \\ 0 & E_0 \left[\frac{1}{2h_t(\beta_0)^2} \frac{\partial h_t(\beta_0)}{\partial \beta} \frac{\partial h_t(\beta_0)}{\partial \beta'} \right] \end{bmatrix}.$$

$$I = \begin{bmatrix} E_0 \left[\frac{1}{h_t(\beta_0)} \frac{\partial m_t(\alpha_0)}{\partial \alpha} \frac{\partial m_t(\alpha_0)}{\partial \alpha'} \right] \\ E_0 \left[\frac{1}{2h_t(\beta_0)^{3/2}} \frac{\partial h_t(\beta_0)}{\partial \beta} \frac{\partial m_t(\alpha_0)}{\partial \alpha'} M_{3t}(\theta_0) \right] \\ E_0 \left[\frac{1}{2h_t(\beta_0)^{3/2}} \frac{\partial m_t(\alpha_0)}{\partial \alpha} \frac{\partial h_t(\beta_0)}{\partial \beta'} M_{3t}(\theta_0) \right] \\ E_0 \left[\frac{1}{4h_t(\beta_0)^2} \frac{\partial h_t(\beta_0)}{\partial \beta} \frac{\partial h_t(\beta_0)}{\partial \beta'} (K_t(\theta_0) - 1) \right] \end{bmatrix}.$$

Nous en déduisons que les deux estimateurs $\hat{\alpha}_T, \hat{\beta}_T$ sont asymptotiquement corrélés dès que la loi conditionnelle des erreurs présente certaines propriétés d'asymétrie. Les matrices de variance-covariance asymptotiques des estimateurs sont :

$$V_{as} \left(\sqrt{T} (\hat{\alpha}_T - \alpha_0) \right) = \left\{ E_0 \left[\frac{1}{h_t(\beta_0)} \frac{\partial m_t(\alpha_0)}{\partial \alpha} \frac{\partial m_t(\alpha_0)}{\partial \alpha'} \right] \right\}^{-1},$$

et

$$\begin{aligned} V_{as} \left(\sqrt{T} (\hat{\beta}_T - \beta_0) \right) &= \left\{ E_0 \left[\frac{1}{2h_t(\beta_0)^2} \frac{\partial h_t(\beta_0)}{\partial \beta} \frac{\partial h_t(\beta_0)}{\partial \beta'} \right] \right\}^{-1} \\ &E_0 \left\{ \frac{1}{4h_t(\beta_0)^2} \frac{\partial h_t(\beta_0)}{\partial \beta} \frac{\partial h_t(\beta_0)}{\partial \beta'} [K_t(\theta_0) - 1] \right\} \\ &\left\{ E_0 \left[\frac{1}{2h_t(\beta_0)^2} \frac{\partial h_t(\beta_0)}{\partial \beta} \frac{\partial h_t(\beta_0)}{\partial \beta'} \right] \right\}^{-1}. \end{aligned}$$

Plus les queues de distributions conditionnelles sont épaisses (au sens kurtosis grande), moins les estimateurs des paramètres figurant dans la variance conditionnelle sont précis.

iii) Modèle de régression avec erreurs ARCH gaussiennes

Dans le cas conditionnellement gaussien, la loi conditionnelle est symétrique et, si le paramètre se décompose en :

$$\theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix},$$

où α n'apparaît que dans la moyenne conditionnelle et β que dans la variance conditionnelle, les deux estimateurs $\hat{\alpha}_T$ et $\hat{\beta}_T$ sont asymptotiquement indépendants. Cette indépendance n'a a priori plus de raison d'être satisfaite si certains paramètres figurent simultanément dans les deux moments. Nous allons cependant voir qu'elle peut continuer à l'être dans certains cas particuliers.

Considérons un modèle de régression avec erreurs ARCH :

$$Y_t = X_t b + \varepsilon_t,$$

où :

$$E \left(\varepsilon_t / \underline{Y}_{t-1}, \underline{X}_t \right) = 0,$$

$$V \left(\varepsilon_t / \underline{Y}_{t-1}, \underline{X}_t \right) = c + a_1 \varepsilon_{t-1}^2 + \dots + a_p \varepsilon_{t-p}^2,$$

et où X_t est fonction du passé de Y_t .

Nous avons $\theta = (b', c, a_1, \dots, a_p)'$, et :

$$\begin{cases} m_t(\theta) = X_t b, \\ h_t(\theta) = c + \sum_{j=1}^p a_j \varepsilon_{t-j}^2 \\ \qquad \qquad = c + \sum_{j=1}^p a_j (Y_{t-j} - X_{t-j} b)^2. \end{cases}$$

Il y a deux types de paramètres :

$$\begin{aligned} \alpha &= b, \\ \beta &= \begin{pmatrix} c \\ a \end{pmatrix}; \end{aligned}$$

le premier groupe apparaît à la fois dans la moyenne et dans la variance alors que le second groupe est spécifique de la variance.

La matrice de variance-covariance asymptotique de l'estimateur est :

$$V_{as} \left[\sqrt{T} \begin{pmatrix} \hat{\alpha}_T - \alpha \\ \hat{\beta}_T - \beta \end{pmatrix} \right] = J^{-1} = \begin{bmatrix} J_{\alpha\alpha} & J_{\alpha\beta} \\ J_{\beta\alpha} & J_{\beta\beta} \end{bmatrix}^{-1},$$

avec :

$$J_{\beta\alpha} = - \mathbb{E}_0 \left\{ \frac{1}{h_t^2} \begin{bmatrix} 1 \\ \varepsilon_{t-1}^2 \\ \vdots \\ \varepsilon_{t-p}^2 \end{bmatrix} \left(\sum_{j=1}^p a_j x_{t-j} \varepsilon_{t-j} \right) \right\}.$$

La fonction sous l'espérance étant antisymétrique dans chacune de composantes ε_{t-j} , $j = 1, \dots, p$, on voit facilement que dans le cas de la normalité $J_{\beta\alpha}$ est nul et $\hat{\alpha}_T$, $\hat{\beta}_T$ sont asymptotiquement indépendants.

4.1.3 Méthode en deux étapes

Dans le cas d'un modèle de régression avec erreurs ARCH, il existe une démarche d'estimation en deux étapes utilisant explicitement les formes linéaires dans les paramètres des moments d'ordre 1 et 2. L'idée est la suivante :

1. On considère le moment d'ordre un et le modèle linéaire associé :

$$Y_t = X_t b + \varepsilon_t.$$

On peut alors régresser par moindres carrés ordinaires Y_t sur X_t , ce qui fournit un estimateur convergent \tilde{b}_T de b . Cet estimateur n'est généralement pas efficace, puisqu'il ne prend pas en compte l'hétéroscédasticité

des erreurs. De cet estimateur, nous déduisons des résidus par moindres carrés ordinaires $\tilde{\varepsilon}_t = Y_t - X_t \tilde{b}_T$.

2. On considère alors le modèle associé au moment conditionnel d'ordre deux :

$$E\left(\varepsilon_t^2 / Y_{t-1}\right) = c + a_1 \varepsilon_{t-1}^2 + \dots + a_p \varepsilon_{t-p}^2,$$

c'est-à-dire :

$$\varepsilon_t^2 = c + a_1 \varepsilon_{t-1}^2 + \dots + a_p \varepsilon_{t-p}^2 + \omega_t,$$

avec $E\left(\omega_t / Y_{t-1}\right) = 0$.

Comme les carrés des erreurs sont inconnus, on ne peut effectuer une régression par moindres carrés ordinaires qu'après avoir remplacé ces erreurs par les résidus associés. On note $\tilde{c}_T, \tilde{a}_{1T}, \dots, \tilde{a}_{pT}$ les estimateurs obtenus en régressant $\tilde{\varepsilon}_t^2$ sur $1, \tilde{\varepsilon}_{t-1}^2, \dots, \tilde{\varepsilon}_{t-p}^2$. Ces estimateurs sont convergents de c, a_1, \dots, a_p .

3. La démarche peut être améliorée en prenant en compte le phénomène d'hétéroscédasticité. Ainsi nous déduisons de l'estimation de seconde étape des approximations des volatilités $\tilde{h}_t = \tilde{c}_T + \tilde{a}_{1T} \tilde{\varepsilon}_{t-1}^2 + \dots + \tilde{a}_{pT} \tilde{\varepsilon}_{t-p}^2$. On peut alors réestimer b en régressant Y_t sur X_t avec les poids $1/\tilde{h}_t$. On obtient des estimateurs $\tilde{\tilde{b}}_T$ de type moindres carrés quasi-généralisés, préférables aux estimateurs initiaux \tilde{b}_T .
4. De même, sous l'hypothèse de normalité conditionnelle le terme d'erreur ω_t du modèle linéaire d'ordre deux est conditionnellement hétéroscédastique, de variance conditionnelle :

$$V\left(\omega_t / Y_{t-1}\right) = 2h_t (\theta_0)^2.$$

On peut donc remplacer les estimateurs \tilde{c}_T, \tilde{a}_T par des estimateurs des moindres carrés quasi-généralisés obtenus en régressant $\tilde{\varepsilon}_t^2$ sur $1, \tilde{\varepsilon}_{t-1}^2, \dots, \tilde{\varepsilon}_{t-p}^2$ avec les poids $1/\tilde{h}_t^2$.

Nous notons $\tilde{\tilde{c}}_T, \tilde{\tilde{a}}_T$ les estimateurs obtenus.

4.1.4 Comparaison des deux approches

i) Les modèles ARCH-M

Il faut d'abord noter que si l'approche par pseudo-maximum de vraisemblance est valable de façon générale, la démarche en deux étapes ne peut être utilisée que pour des modèles particuliers. Ainsi si la volatilité intervenait parmi les variables explicatives, par exemple si :

$$Y_t = X_t b + h_t b^* + \varepsilon_t,$$

l'une des variables explicatives h_t serait inobservable. La régression de première étape ne pourrait être menée qu'après avoir remplacé h_t par une approximation \hat{h}_t , c'est-à-dire après avoir effectué l'estimation de deuxième étape. Ainsi dans le cas des modèles ARCH-M, on ne peut estimer de façon séquentielle les paramètres b, b^* et les paramètres c, a .

ii) Efficacité relative

Dans le cas favorable de modèles de régression avec erreur ARCH (ou GARCH) les deux approches peuvent être utilisées. La démarche en deux étapes est la plus simple, mais peut se révéler peu efficace.

Ainsi pour le modèle :

$$\begin{cases} Y_t = b + \varepsilon_t, \\ E(\varepsilon_t / Y_{t-1}) = 0, \\ V(\varepsilon_t / Y_{t-1}) = c + a_1 \varepsilon_{t-1}^2 + \dots + a_p \varepsilon_{t-p}^2, \end{cases} \quad (4.4)$$

il est possible d'établir le résultat suivant.

Propriété 4.2 (Engle [1982], Gouriéroux [1992] chap. IV) :

Pour le modèle (4.4) avec erreurs conditionnellement gaussiennes :

1. l'estimateur \tilde{c}_T, \tilde{a}_T est asymptotiquement efficace ;
2. les estimateurs $(\tilde{c}_T, \tilde{a}_T)'$ et \tilde{b}_T sont asymptotiquement indépendants ;
3. l'estimateur \tilde{b}_T n'est pas asymptotiquement efficace.

De plus cette perte d'efficacité peut être précisée en examinant l'efficacité asymptotique relative :

$$EAR = \lim_T \frac{V_{as} \left(\sqrt{T} (\tilde{b}_T - b_0) \right)}{V_{as} \left(\sqrt{T} (\hat{b}_T - b_0) \right)}.$$

On peut vérifier que cette quantité tend vers l'infini, lorsque les vraies valeurs des paramètres sont telles que $a_1 + \dots + a_p$ tende vers un. Ainsi il y a une très importante perte de précision au voisinage de l'absence de stationnarité faible, cas limite qui est comme nous l'avons vu fréquemment rencontré en pratique.

4.1.5 Méthode des moments généralisés

Le modèle hétéroscédastique est en fait défini par l'intermédiaire de conditions d'orthogonalité. Les contraintes :

$$\begin{cases} \mathbb{E}_0 \left[Y_t / \underline{Y}_{t-1} \right] = m_t(\theta_0), \\ \mathbb{V}_0 \left[Y_t / \underline{Y}_{t-1} \right] = h_t(\theta_0), \end{cases}$$

sont en effet équivalentes à :

$$\begin{cases} \mathbb{E}_0 \left[Y_t / \underline{Y}_{t-1} \right] = m_t(\theta_0), \\ \mathbb{E}_0 \left[Y_t^2 / \underline{Y}_{t-1} \right] = h_t(\theta_0) + m_t(\theta_0)^2. \end{cases}$$

Elles peuvent alors s'écrire :

$$\begin{cases} \mathbb{E}_0 \left(\tilde{Z}_t (Y_t - m_t(\theta_0)) \right) = 0, \\ \mathbb{E}_0 \left(\tilde{Z}_t (Y_t^2 - h_t(\theta_0) - m_t(\theta_0)^2) \right) = 0, \end{cases}$$

pour toute variable \tilde{Z}_t (de carré intégrable) fonction des valeurs passées de Y . Une telle variable Z_t est appelée variable instrumentale ou instrument.

Notons

$$\rho(Y_t, \theta) = \begin{bmatrix} Y_t - m_t(\theta) \\ Y_t^2 - h_t(\theta) - m_t(\theta)^2 \end{bmatrix},$$

et introduisons une matrice Z_t de variables instrumentales, de taille $(k, 2)$ où k désigne le nombre de paramètres figurant dans θ . Nous avons les conditions d'orthogonalité :

$$\mathbb{E}_0 [Z_t \rho(Y_t, \theta_0)] = 0.$$

On peut alors estimer la vraie valeur inconnue θ_0 en cherchant une solution de la condition d'orthogonalité approchée :

$$\sum_{t=1}^T Z_t \rho(Y_t, \tilde{\theta}_T(Z)) = 0. \quad (4.5)$$

Le système ayant autant d'équations que d'inconnues permet en général une résolution non ambiguë. La solution obtenue est dite estimateur à variables instrumentales. Son expression et ses propriétés asymptotiques dépendent des instruments (Z_t) retenus. Sous des conditions de régularité usuelles (voir Hansen [1982], Chamberlain [1987], Newey [1989]), on montre que l'estimateur $\tilde{\theta}_T(Z)$ est convergent, asymptotiquement normal, et de matrice de variance-covariance

asymptotique donnée par :

$$\begin{aligned} \Lambda(Z) &= E_0 \left[Z_t \frac{\partial \rho}{\partial \theta'} (Y_t, \theta_0) \right]^{-1} \lim_T V_0 \left[\frac{1}{\sqrt{T}} \sum Z_t \rho (Y_t, \theta_0) \right] \\ &\quad E_0 \left[\frac{\partial \rho'}{\partial \theta} (Y_t, \theta_0) Z_t' \right]^{-1} \\ &= E_0 \left[Z_t \frac{\partial \rho}{\partial \theta'} (Y_t, \theta_0) \right]^{-1} E_0 [Z_t \rho (Y_t, \theta_0) \rho (Y_t, \theta_0)' Z_t'] \\ &\quad E_0 \left[\frac{\partial \rho'}{\partial \theta} (Y_t, \theta_0) Z_t' \right]^{-1}, \end{aligned}$$

en utilisant le fait que :

$$E_0 \left[\rho (Y_t, \theta_0) / \underline{Y}_{t-1} \right] = 0$$

et que Z_t est fonction du passé de Y . Cette matrice peut encore s'écrire :

$$\begin{aligned} \Lambda(Z) &= E_0 \left[Z_t E_0 \left(\frac{\partial \rho}{\partial \theta'} (Y_t, \theta_0) / \underline{Y}_{t-1} \right) \right]^{-1} \\ &\quad E_0 \left(Z_t E_0 \left[\rho (Y_t, \theta_0) \rho (Y_t, \theta_0)' / \underline{Y}_{t-1} \right] Z_t' \right) \quad (4.6) \\ &\quad E_0 \left[E_0 \left(\frac{\partial \rho'}{\partial \theta} (Y_t, \theta_0) / \underline{Y}_{t-1} \right) Z_t' \right]^{-1}. \end{aligned}$$

Parmi cette classe d'estimateurs à variables instrumentales, on vérifie facilement que le choix

$$Z_t^* = E_0 \left[\frac{\partial \rho'}{\partial \theta} (Y_t, \theta_0) / \underline{Y}_{t-1} \right] V_0 \left[\rho (Y_t, \theta_0) / \underline{Y}_{t-1} \right]^{-1} \quad (4.7)$$

conduit à un estimateur asymptotiquement préférable à tous les autres. Un tel estimateur est dit des moments généralisés. Sa précision asymptotique vaut :

$$\begin{aligned} \Lambda(Z^*) &= \left\{ E_0 \left(\frac{\partial \rho'}{\partial \theta} (Y_t, \theta_0) / \underline{Y}_{t-1} \right) V_0 \left(\rho (Y_t, \theta_0) / \underline{Y}_{t-1} \right)^{-1} \right. \\ &\quad \left. E_0 \left(\frac{\partial \rho}{\partial \theta'} (Y_t, \theta_0) / \underline{Y}_{t-1} \right) \right\}^{-1}, \end{aligned} \quad (4.8)$$

et est l'inverse de la borne d'efficacité semi-paramétrique associée aux conditions de moments

$$E_0 \left[\rho (Y_t, \theta_0) / \underline{Y}_{t-1} \right] = 0 \text{ (Chamberlain [1987]).}$$

En pratique évidemment les instruments optimaux Z_t^* dépendent des vraies lois inconnues et doivent être remplacés par des approximations convergentes afin de permettre le calcul d'un estimateur de type moment généralisé.

Dans notre cas particulier, nous avons :

$$E_0 \left[\frac{\partial \rho}{\partial \theta'} (Y_t, \theta_0) / \underline{Y_{t-1}} \right] = \begin{bmatrix} -\frac{\partial m_t}{\partial \theta'} (\theta_0) \\ -\frac{\partial h_t}{\partial \theta'} (\theta_0) - 2m_t (\theta_0) \frac{\partial m_t}{\partial \theta'} (\theta_0) \end{bmatrix},$$

$$V_0 \left[\rho (Y_t, \theta_0) / \underline{Y_{t-1}} \right]$$

$$= \begin{pmatrix} V_0 \left[Y_t / \underline{Y_{t-1}} \right] & Cov_0 \left[Y_t, Y_t^2 / \underline{Y_{t-1}} \right] \\ Cov_0 \left[Y_t, Y_t^2 / \underline{Y_{t-1}} \right] & V_0 \left[Y_t^2 / \underline{Y_{t-1}} \right] \end{pmatrix}$$

$$= \begin{bmatrix} h_t (\theta_0) & h_t (\theta_0)^{3/2} M_{3t} (\theta_0) + 2m_t (\theta_0) h_t^2 (\theta_0) \\ - & h_t^2 (\theta_0) K_t (\theta_0) + 4m_t^2 (\theta_0) h_t^2 (\theta_0) \\ & + 4m_t (\theta_0) h_t^{3/2} (\theta_0) M_{3t} (\theta_0) \end{bmatrix}.$$

4.2 PREVISION

4.2.1 Forme des intervalles de prévision

Considérons un modèle ARMA avec erreurs GARCH conditionnellement gaussiennes :

$$\begin{cases} \Phi(L)Y_t = \Theta(L)\varepsilon_t \\ \text{où } \varepsilon_t / \underline{\varepsilon_{t-1}} \sim N(0, h_t), \\ h_t = c + \sum_{i=1}^p a_i \varepsilon_{t-i}^2 + \sum_{j=1}^q b_j h_{t-j}. \end{cases}$$

Un tel modèle peut être analysé de deux façons différentes :

1. On peut dans un premier temps appliquer les procédures classiques d'estimation et d'analyse des processus ARMA, c'est-à-dire faire comme si les données étaient conditionnellement homoscédastiques. Les estimateurs des coefficients des polynômes Φ et Θ , qui ne sont autres que les estimateurs de première étape de la méthode en deux étapes décrite en 4.1. sont convergents. Les prévisions à horizon 1 des Y_t , c'est-à-dire les variables

$$\hat{Y}_t = \left[\frac{\hat{\Phi}(L)}{\hat{\Theta}(L)} - 1 \right] Y_t,$$

sont asymptotiquement sans biais. Dans cette démarche, la variabilité est estimée par :

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (Y_t - \hat{Y}_t)^2 = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^2,$$

et des intervalles de prévisions sont :

$$[\hat{Y}_t \pm 2\hat{\sigma}]$$

(en négligeant l'effet d'estimation de Φ et Θ).

$\hat{\sigma}^2$ n'est autre qu'une estimation convergente de :

$$\begin{aligned} E \left[Y_t - E(Y_t / \underline{Y}_{t-1}) \right]^2 &= E E \left[\left[Y_t - E(Y_t / \underline{Y}_{t-1}) \right]^2 / \underline{Y}_{t-1} \right] \\ &= E V(Y_t / \underline{Y}_{t-1}) \\ &= E h_t, \end{aligned}$$

c'est-à-dire de la valeur moyenne de la volatilité. Elle est en particulier indépendante de la date t de prévision, de sorte que tous les intervalles de prévision ont la même largeur.

2. Dans un deuxième temps, on peut tenir compte du modèle d'évolution de la volatilité et appliquer les procédures d'estimation spécifiques aux modèles ARCH. Si $\hat{\hat{\Phi}}$, $\hat{\hat{\Theta}}$ désignent les estimateurs des polynômes autorégressif et moyenne mobile, les prévisions à horizon 1 des Y_t , données par :

$$\hat{\hat{Y}}_t = \left[\frac{\hat{\hat{\Phi}}(L)}{\hat{\hat{\Theta}}(L)} - 1 \right] Y_t,$$

sont asymptotiquement sans biais. Les intervalles de prévision sont maintenant calculés par :

$$[\hat{\hat{Y}}_t \pm 2\hat{h}_t],$$

où \hat{h}_t est l'estimation de la volatilité de la date t . La largeur de ces intervalles dépend maintenant de la date t considérée.

4.2.2 Un exemple

A titre d'illustration, nous considérons l'analyse du taux d'inflation américain π_t sur données trimestrielles de 1948.2 à 1983.4, soit un total de 143 observations (voir Engle-Kraft [1983] et Bollerslev [1986]).

1. L'utilisation des approches standards de séries temporelles conduit à l'identification du modèle :

$$\left\{ \begin{array}{l} \pi_t = 0.240 + 0.552 \pi_{t-1} + 0.177 \pi_{t-2} + 0.232 \pi_{t-3} \\ \quad (3.00) \quad (6.65) \quad (1.99) \quad (2.58) \\ \quad - 0.209 \pi_{t-4} + \varepsilon_t, \\ \quad (2.61) \\ h_t = 0.282, \\ \quad (8.29) \end{array} \right. \quad (4.9)$$

et à des prévisions et intervalles de prévision décrits dans la figure 4.1 :

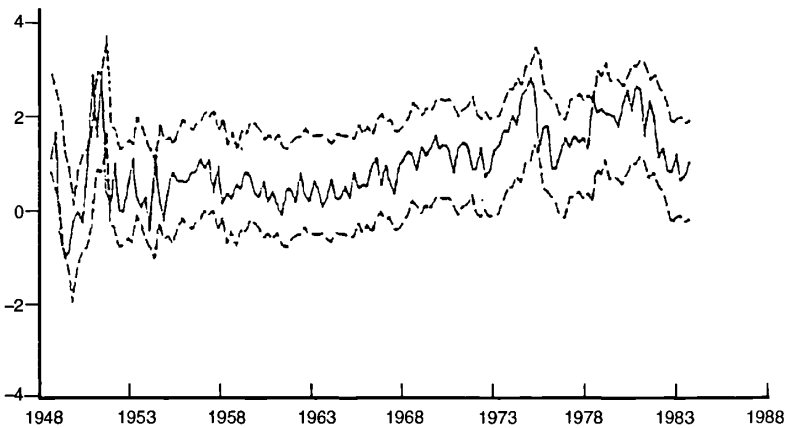


Figure 4.1: Intervalles de prévision pour l'estimation ARMA

2. L'utilisation d'un modèle ARMA-GARCH permet de mettre en évidence un effet d'hétéroscédasticité. La spécification estimée est :

$$\left\{ \begin{array}{l} \pi_t = 0.141 + 0.433 \pi_{t-1} + 0.229 \pi_{t-2} + 0.349 \pi_{t-3} \\ \quad (2.35) \quad (5.34) \quad (2.08) \quad (4.53) \\ \quad - 0.162 \pi_{t-4} + \varepsilon_t, \\ \quad (1.56) \\ h_t = 0.007 + 0.135 \varepsilon_{t-1}^2 + 0.829 h_{t-1}. \\ \quad (1.17) \quad (1.93) \quad (12.19) \end{array} \right. \quad (4.10)$$

Les intervalles de prévision sont de largeur plus grande au moment où les prévisions sont plus variables (voir figure 4.2).

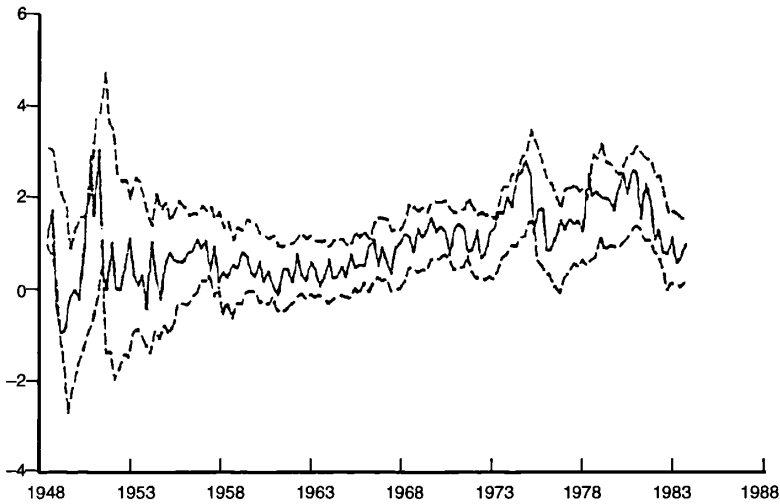


Figure 4.2: Intervalles de prévision pour l'estimation ARMA-GARCH

On notera que

- (a) les centres des intervalles de prévision sont approximativement les mêmes avec les deux approches ;
- (b) la valeur moyenne de la volatilité pour le modèle ARMA-GARCH vaut :

$$Eh_t = \frac{0.007}{1 - 0.135 - 0.829} = 0.195,$$

et est non significativement différente de la valeur 0.282 déduite de la démarche standard ;

- (c) les coefficients résumant la dynamique de la volatilité sont tels que :

$$0.135 + 0.829 = 0.964.$$

La somme est proche de la valeur limite 1 correspondant à la non-existence des moments d'ordre deux.

4.2.3 Volatilité et stratégies d'intervention

Une justification usuelle des modèles ARMA est la possibilité d'obtenir des prévisions sans biais de toute fonction linéaire des observations (Y_t), même si ces prévisions ne sont pas forcément très précises. Cette propriété importante se révèle cependant insuffisante pour les applications financières, où la notion de risque doit être prise en compte. Or ce risque est lié à la volatilité, donc à la prévision d'une fonction quadratique des observations. Dans l'exemple précédent, nous avons vu des différences significatives pour la construction des

intervalles de prévision, selon la démarche ARMA ou ARMA-GARCH utilisée. Ces différences ont des conséquences importantes au niveau financier, comme nous allons le voir à partir d'une stratégie simple d'intervention sur le marché.

Nous considérons un investisseur constituant son portefeuille à partir d'un actif sans risque et d'un actif risqué. Nous supposons que l'actif sans risque a un rendement fixe r_f indépendant du temps. L'actif risqué a un rendement r_t aléatoire pour lequel on dispose d'un intervalle de prévision à horizon 1 $[\hat{r}_t \pm 2\hat{h}_t]$. Une stratégie simple d'achat et de vente est alors la suivante :

- L'investisseur disposant dans son portefeuille d'actif sans risque l'échange contre de l'actif risqué, à la première date pour laquelle :

$$\hat{r}_t - 2\hat{h}_t \geq r_f.$$

- L'investisseur disposant dans son portefeuille d'actif risqué l'échange contre de l'actif sans risque à la première date pour laquelle :

$$\hat{r}_t - 2\hat{h}_t \leq r_f.$$

A chaque date le portefeuille est homogène, comprenant soit uniquement de l'actif risqué, soit uniquement de l'actif sans risque. De façon claire la stratégie d'intervention et le gain financier résultant de cette stratégie dépendent beaucoup de la façon de déterminer les intervalles de prévision. Nous décrivons ci-dessous une telle stratégie, où les instants d'achat (A) et de vente (V) sont indiqués pour le modèle hétéroscédastique.

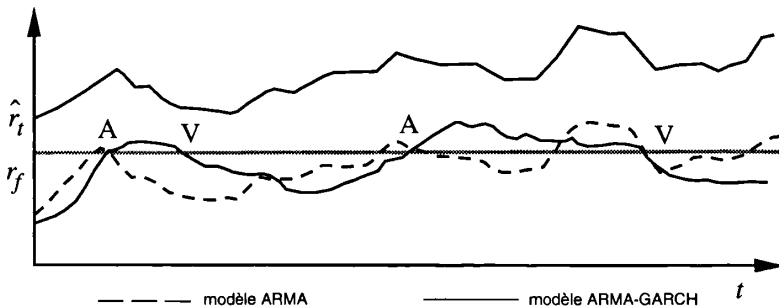


Figure 4.3

Stratégies d'intervention fondées sur un modèle ARMA et sur un modèle ARMA-GARCH

4.3 TESTS D'HYPOTHÈSES

Des procédures de test peuvent être développées selon les approches usuelles à partir des méthodes d'estimation décrites en 4.1. Dans cette section-ci

nous n'insisterons pas sur cet aspect. Nous allons plutôt étudier de façon détaillée deux types d'hypothèses importantes pour les applications financières :

1. L'hypothèse de « marche aléatoire », qui porte sur la forme du moment conditionnel d'ordre un, mais dont l'analyse doit être modifiée en présence d'hétéroscédasticité conditionnelle ;
2. L'hypothèse d'homoscédasticité conditionnelle, qui porte sur le moment conditionnel d'ordre deux et dont le test devrait précéder tout passage d'un modèle ARMA standard à un modèle plus complexe de type ARMA-GARCH.

4.3.1 Hypothèse de marche aléatoire

i) Importance de l'hypothèse en finance

La propriété de marche aléatoire (ou de martingale) s'introduit naturellement et pour des raisons assez diverses dans l'analyse des données financières. Nous décrivons ci-dessous certains des arguments classiques.

a) Argument d'absence d'opportunité d'arbitrage

Considérons deux actifs : l'un non risqué, de rendement fixe déterministe r_f , l'autre risqué de rendement aléatoire r_t . Soit P_t le prix de l'actif risqué à la date t . Nous avons :

$$P_{t+T} = P_t (1 + r_{t+1}) \dots (1 + r_{t+T}).$$

Supposons que la suite des prix soit telle que :

$$\forall T, E \left[P_{t+T} / \underline{P_{t+T-1}} \right] > (1 + r_f) P_{t+T-1},$$

c'est-à-dire que la suite corrigée du rendement de l'actif non risqué

$$\left[P_{t+T} / (1 + r_f)^T \right]$$

constitue une sous-martingale. Un investisseur conservant pendant un temps T très long de l'actif risqué ferait un gain positif certain par rapport à un investisseur conservant pendant un temps T très long de l'actif sans risque. Des investisseurs employant des stratégies de ce type sont alors conduits à ne retenir que de l'actif risqué ; s'il n'y a que de tels investisseurs sur le marché, il n'y aura plus de demande d'actif non risqué et le marché de celui-ci disparaîtra.

Un raisonnement symétrique conduit à rejeter la possibilité que le processus constitue une surmartingale

$$\forall T, E \left[P_{t+T} / \underline{P_{t+T-1}} \right] < (1 + r_f) P_{t+T-1}.$$

Ceci conduit à faire jouer un rôle particulier à la condition limite :

$$\forall T, E \left[P_{t+T} / \underline{P_{t+T-1}} \right] = (1 + r_f) P_{t+T-1}.$$

On notera cependant :

- que tout processus pour lequel

$$E \left[P_{t+T} / \underline{P_{t+T-1}} \right] > (1 + r_f) P_{t+T-1}$$

pour certaines valeurs de T , et

$$E \left[P_{t+T} / \underline{P_{t+T-1}} \right] < (1 + r_f) P_{t+T-1}$$

pour d'autres, serait compatible avec une absence d'opportunité d'arbitrage fondée sur des stratégies longues,

- que les investisseurs peuvent avoir des stratégies plus complexes, dans lesquelles l'aspect risque est plus important,
- qu'il est alors naturel qu'existe une prime de risque, c'est-à-dire que :

$$E \left[P_{t+T} / \underline{P_{t+T-1}} \right] > (1 + r_f) P_{t+T-1}.$$

b) Argument de comportement des intervenants

Cet argument est du type suivant : si un investisseur pense que $E[P_{t+T}/P_t]$ est « sensiblement supérieur » à $(1 + r_f)^T P_t$, il aura tendance à acheter de l'actif risqué à la date t ; il y aura alors accroissement à cette date de la demande de cet actif, donc augmentation du prix P_t et diminution de l'écart

$$E \left[P_{t+T} / \underline{P_t} \right] - [1 + r_f]^T P_t.$$

Par leurs interventions les investisseurs ont tendance à rapprocher l'évolution des prix corrigés du sans risque de celle d'une martingale.

c) Argument descriptif

En conséquence, on s'attend à voir approximativement satisfaite cette propriété de martingale pour les marchés très liquides, où de grandes quantités sont régulièrement échangées. Les tests usuels de marche aléatoire appliqués à de telles données semblent confirmer partiellement ce phénomène. On prendra garde cependant que ces tests souvent effectués sans tenir compte de l'hétéroscédasticité conditionnelle peuvent être biaisés, que les résultats des tests peuvent dépendre sensiblement du mode de calcul du rendement (façon d'intégrer les dividendes, les taux d'imposition, les conséquences d'attribution d'actions gratuites, les effets de week-end), de la définition retenue pour l'actif (nous renvoyons ici à l'exemple du chapitre 3 concernant l'indice de marché), de l'unité de temps retenue (données

journalières, hebdomadaires, mensuelles), que l'hypothèse n'a de chances d'être satisfaite que pour des marchés constamment très liquides, ce qui élimine les titres où apparaît à un moment donné un investisseur prépondérant (titres ayant été nationalisés, ou soumis à OPA).

d) Argument d'équilibre

L'écriture de modèles d'équilibre fixant les évolutions de prix et de quantités (voir chapitre 7) montrent que l'écart :

$$E \left[P_{t+T} / P_{t+T-1} \right] - (1 + r_f) P_{t+T-1}$$

est fonction du risque et de l'importance que les investisseurs attribuent à ce risque. Dans le cas limite, où ces investisseurs n'ont pas peur du risque (neutralité vis-à-vis du risque), cette différence doit être nulle à l'équilibre et donc l'hypothèse de martingale être satisfaite. Cette interprétation est plus intéressante car elle montre que plus que l'hypothèse de martingale, c'est l'importance de l'écart à cette hypothèse dont il sera utile d'étudier l'évolution et en particulier de la lier à celle du risque.

e) Argument de valorisation

Dans le cas général, on peut établir en invoquant des arguments d'arbitrage que les prix des divers actifs peuvent être exprimés en fonction des prix d'actifs de base que sont les actifs contingents : brièvement un actif contingent est un actif qui donne un franc en $t + T$, si certaines conditions se réalisent. De cette interprétation on déduit que les prix des actifs constituent des martingales à condition de raisonner avec une loi de probabilité modifiée prenant en compte la loi initiale et l'évolution de ces prix d'actifs contingents. Tester l'hypothèse de marche aléatoire revient alors à tester s'il est ou non nécessaire d'effectuer un changement de loi de probabilité. Ce point est crucial pour valoriser de nouveaux actifs, en particulier des options plus ou moins exotiques.

ii) La procédure de test classique

L'analyse est en pratique effectuée directement sur les rendements, et non sur les prix, de sorte que l'hypothèse de marche aléatoire qui du point de vue financier porte sur les prix (ou les logarithmes des prix) se ramène à une hypothèse de bruit blanc sur les rendements. Notons $(Y_t) = (\varepsilon_t)$ cette série. La procédure standard est le test portmanteau. Celui-ci consiste à calculer les autocorrélations :

$$\hat{\rho}_T(h) = \frac{\sum_{t=1}^{T-h} \varepsilon_t \varepsilon_{t-h}}{\sum_{t=1}^T \varepsilon_t^2},$$

puis à déterminer la statistique portmanteau :

$$Q_H = T \sum_{h=1}^H \hat{\rho}_T(h)^2. \quad (4.11)$$

Si $(Y_t) = (\varepsilon_t)$ est un bruit blanc indépendant, on établit que la statistique Q_H suit asymptotiquement une loi du khi-deux à H degrés de liberté. Si $\chi_{95\%}^2(H)$ désigne le quantile à 95 % de cette loi, le test consiste :

$$\left\{ \begin{array}{l} \text{à accepter l'hypothèse de bruit blanc, si} \\ \quad \quad \quad Q_H < \chi_{95\%}^2(H), \\ \text{à la refuser sinon.} \end{array} \right. \quad (4.12)$$

Nous donnons ci-dessous les résultats de tels tests effectués sur séries américaines, période Janvier 1966-Décembre 1976 (Taylor [1986]). H est égal à 10 et la valeur critique est 18.31.

Tableau 4.1: Test de bruit blanc

Série	Valeur de la statistique
Bon du Trésor	11.39
Taux de change DM/\$ à 6 mois	29.76
Blé à 12 mois	15.51
Blé à 6 mois	22.04
Blé à 3 mois	27.15
Or	29.31
General Motors	25.89
Kodak	19.91
Dupont	52.04

iii) Test portmanteau corrigé

La forme retenue de la statistique Q_H et celle de sa loi asymptotique reposent sur l'hypothèse d'indépendance du bruit blanc. Lorsque le bruit blanc n'est pas indépendant, notamment lorsqu'il présente de l'hétéroscédasticité conditionnelle, la procédure précédente est biaisée. Nous allons expliquer comment corriger la statistique portmanteau. Rappelons que dans le cas standard, la procédure portmanteau est équivalente à un test du multiplicateur de Lagrange de l'hypothèse :

$$H_0 = \{a_1 = \dots = a_H = 0\},$$

dans la représentation autorégressive :

$$\varepsilon_t = a_1 \varepsilon_{t-1} + \dots + a_H \varepsilon_{t-H} + u_t \quad t = 1, \dots, T - H.$$

C'est sous cette forme que nous allons construire la statistique de test. Les estimateurs des moindres carrés ordinaires des coefficients autorégressifs sont :

$$\begin{aligned} \begin{bmatrix} \hat{a}_1 \\ \vdots \\ \hat{a}_H \end{bmatrix} &= \begin{pmatrix} \Sigma \varepsilon_{t-1}^2 & \dots & \Sigma \varepsilon_{t-1} \varepsilon_{t-H} \\ \vdots & & \vdots \\ \Sigma \varepsilon_{t-1} \varepsilon_{t-H} & \dots & \Sigma \varepsilon_{t-H}^2 \end{pmatrix}^{-1} \begin{pmatrix} \Sigma \varepsilon_t \varepsilon_{t-1} \\ \vdots \\ \Sigma \varepsilon_t \varepsilon_{t-H} \end{pmatrix} \\ &= \begin{pmatrix} \Sigma \varepsilon_{t-1}^2 & \dots & \Sigma \varepsilon_{t-1} \varepsilon_{t-H} \\ \vdots & & \vdots \\ \Sigma \varepsilon_{t-1} \varepsilon_{t-H} & \dots & \Sigma \varepsilon_{t-H}^2 \end{pmatrix}^{-1} T \begin{bmatrix} \hat{\gamma}_T(1) \\ \vdots \\ \hat{\gamma}_T(H) \end{bmatrix}, \end{aligned}$$

où $\hat{\gamma}_T(h)$ est l'autocovariance empirique d'ordre h .

Si nous tenons compte de l'hétéroscédasticité conditionnelle éventuelle, nous avons sous l'hypothèse nulle :

$$\sqrt{T} \begin{bmatrix} \hat{\gamma}_T(1) \\ \vdots \\ \hat{\gamma}_T(H) \end{bmatrix} \xrightarrow{\text{loi}} N[0, \Omega],$$

où la matrice de variance asymptotique se déduit directement de celle de l'estimateur m.c.o. en présence d'hétéroscédasticité. Cette matrice est donnée par :

$$\begin{aligned} \Omega &= \lim_{T \rightarrow \infty} \frac{1}{T-h} \sum_{t=1}^{T-h} \begin{bmatrix} \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-H} \end{bmatrix} h_t [\varepsilon_{t-1}, \dots, \varepsilon_{t-H}] \\ &= E \left\{ \begin{bmatrix} \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-H} \end{bmatrix} h_t [\varepsilon_{t-1}, \dots, \varepsilon_{t-H}] \right\} \\ &= E \left\{ \begin{bmatrix} \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-H} \end{bmatrix} E \left(\varepsilon_t^2 / \varepsilon_{t-1} \right) (\varepsilon_{t-1}, \dots, \varepsilon_{t-H}) \right\} \\ &= E \left[\begin{bmatrix} \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-H} \end{bmatrix} \varepsilon_t^2 (\varepsilon_{t-1}, \dots, \varepsilon_{t-H}) \right]. \end{aligned}$$

Cette matrice est estimée de façon convergente par :

$$\hat{\Omega}_T = \frac{1}{T-H} \sum_{t=1}^{T-H} \begin{bmatrix} \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-H} \end{bmatrix} \varepsilon_t^2 [\varepsilon_{t-1}, \dots, \varepsilon_{t-H}]. \tag{4.13}$$

La statistique portmanteau modifiée pour tenir compte de l'hétéroscédasti-

cit  est :

$$\tilde{Q}_H = T [\hat{\gamma}_T(1), \dots, \hat{\gamma}_T(H)] \hat{\Omega}_T^{-1} \begin{bmatrix} \hat{\gamma}_T(1) \\ \vdots \\ \hat{\gamma}_T(H) \end{bmatrix} \quad (4.14)$$

Elle suit asymptotiquement sous l'hypoth se nulle la loi $\chi^2(H)$ et la proc dure de test consiste   accepter cette hypoth se si $\hat{Q}_H < \chi_{95\%}^2(H)$,   la refuser sinon.

Compar e   la statistique initiale Q_H , la nouvelle statistique attribue des poids diff rents aux p riodes de forte volatilit  [h_t ou ε_t^2 grand] et de petite volatilit  [h_t ou ε_t^2 petit]. Il s'agit de la correction de White [1980] usuelle.

Dans le cas particulier conditionnellement homoscedastique, nous avons :

$$\begin{aligned} \Omega &= E \left[\begin{pmatrix} \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-H} \end{pmatrix} E \left(\varepsilon_t^2 / \varepsilon_{t-1} \right) (\varepsilon_{t-1}, \dots, \varepsilon_{t-H}) \right] \\ &= E (\varepsilon_t^2) E \left[\begin{pmatrix} \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-H} \end{pmatrix} (\varepsilon_{t-1}, \dots, \varepsilon_{t-H}) \right] \\ &= \gamma(0)^2 Id. \end{aligned}$$

La statistique peut  tre prise  gale   :

$$\begin{aligned} \tilde{Q}_H &= T [\hat{\gamma}_T(1), \dots, \hat{\gamma}_T(H)] \frac{Id}{\hat{\gamma}_T(0)^2} \begin{pmatrix} \hat{\gamma}_T(1) \\ \vdots \\ \hat{\gamma}_T(H) \end{pmatrix} \\ &= T \sum_{h=1}^H \hat{\rho}_T(h)^2, \end{aligned}$$

c'est- -dire   la statistique portmanteau usuelle.

4.3.2 Test d'homoscedasticit 

Consid rons un mod le de r gression avec erreurs h t roscedastiques. Nous supposons que les moyenne et variance conditionnelles d pendent de param tres $\theta = (\alpha', \beta_0, \beta_1)'$ de la fa on suivante :

$$\begin{cases} m_t(\theta) &= m_t(\alpha), \\ h_t(\theta) &= h_t(\alpha, \beta_0, \beta_1), \end{cases} \quad (4.15)$$

et que $h_t(\alpha, \beta_0, 0) = h(\beta_0)$.

Ainsi certains param tres $\beta = (\beta_0, \beta_1)'$ sont sp cifiques de la variance et lorsque les sous-param tres β_1 sont nuls le mod le est conditionnellement homoscedastique.

Exemple :

Considérons ainsi un modèle de régression avec erreurs ARCH.

$$\begin{cases} Y_t = X_t b + \varepsilon_t, \\ E(\varepsilon_t / \underline{Y}_{t-1}, \underline{X}_t) = 0, \\ V(\varepsilon_t / \underline{Y}_{t-1}, \underline{X}_t) = c + a_1 \varepsilon_{t-1}^2 + \dots + a_p \varepsilon_{t-p}^2. \end{cases}$$

Nous avons :

$$h_t(\theta) = c + a_1 (Y_{t-1} - X_{t-1} b)^2 + \dots + a_p (Y_{t-p} - X_{t-p} b)^2,$$

et la forme précédente est satisfaite avec :

$$\alpha = b, \beta_0 = c, \beta_1 = (a_1, \dots, a_p)'$$

Lorsqu'il y a homoscedasticité, c'est-à-dire sous l'hypothèse

$$H_0 = \{a_1 = \dots = a_p = 0\},$$

le paramètre b n'apparaît plus dans l'expression de la variance conditionnelle.

Lorsque le modèle est conditionnellement gaussien, le test du multiplicateur de Lagrange de l'hypothèse d'homoscedasticité conditionnelle :

$$H_0 = \{\beta_1 = 0\},$$

conduit à employer la démarche intuitive suivante (voir Gouriéroux [1992] chap.IV.D.), fondée sur le développement limité de l'équation de volatilité. Nous avons :

$$\varepsilon_t^2 = h_t(\alpha, \beta_0, \beta_1) + u_t,$$

avec $E(u_t / \underline{u}_{t-1}) = 0, V(u_t / \underline{u}_{t-1}) = 2h_t^2(\alpha, \beta_0, \beta_1)$.

Ce modèle peut être approché au voisinage de l'hypothèse nulle $H_0 = (\beta_1 = 0)$ par :

$$\begin{cases} \varepsilon_t^2 \simeq h(\beta_0) + \frac{\partial h_t}{\partial \beta_1'}(\alpha, \beta_0, 0) \beta_1 + u_t, \\ E(u_t / \underline{u}_{t-1}) = 0, V(u_t / \underline{u}_{t-1}) = 2h^2(\beta_0). \end{cases}$$

Il peut aussi l'être en remplaçant les paramètres inconnus par leurs estimateurs sous l'hypothèse nulle ; introduisons les estimateurs par m.c.o. de α et β_0 , notés respectivement $\hat{\alpha}^0$ et $\hat{\beta}_0^0$. Ils sont définis par :

$$\begin{aligned} \hat{\alpha}^0 &= \text{Arg min}_{\alpha} \sum_{t=1}^T [Y_t - m_t(\alpha)]^2, \\ h(\hat{\beta}_0^0) &= \frac{1}{T} \sum_{t=1}^T [Y_t - m_t(\hat{\alpha}^0)]^2, \end{aligned}$$

et notons : $\hat{\varepsilon}_t^0 = Y_t - m_t(\hat{\alpha}^0)$ les résidus m.c.o. associées. Le modèle peut aussi être approché par :

$$\begin{cases} \hat{\varepsilon}_t^0 \simeq h(\hat{\beta}_0) + \frac{\partial h_t}{\partial \beta_1'}(\hat{\alpha}^0, \hat{\beta}_0^0, 0)\beta_1 + u_t, \\ E(u_t/u_{t-1}) = 0, V(u_t/u_{t-1}) = 2h^2(\hat{\beta}_0^0). \end{cases} \quad (4.16)$$

On établit alors le résultat suivant :

Propriété 4.3

La statistique du multiplicateur de Lagrange de l'hypothèse H_0 d'homoscédasticité conditionnelle est :

$$\xi = TR^2,$$

où R^2 est le coefficient de corrélation multiple dans le modèle homoscédastique approché (4.16). Cette statistique suit asymptotiquement sous l'hypothèse nulle une loi du khi-deux avec un nombre de degrés de liberté égal à la dimension du vecteur β_1 .

4.4 DU GENERAL AU SPECIFIQUE

4.4.1 Le principe

Nous avons vu dans le chapitre 3 qu'il existe diverses spécifications paramétriques possibles pour la dynamique de la volatilité. Choisir a priori l'une d'entre elles peut conduire à des erreurs de spécification et à des biais sur l'évaluation du risque. Une démarche préférable consiste :

1. à commencer par spécifier de façon adéquate la forme de la moyenne conditionnelle par les procédures usuelles. On en déduit alors des résidus $\hat{\varepsilon}_t^2$, de type résidus m.c.o. ;
2. à étudier de façon non paramétrique la régression de $\hat{\varepsilon}_t^2$ sur le passé $\hat{\varepsilon}_{t-1}$, soit par des approches de type noyau (voir Gallant-Tauchen [1989], Pagan-Schwert [1989]), soit en utilisant des bases adéquates de fonctions dépendant du passé ;
3. finalement à retenir une formulation paramétrique compatible avec l'estimation non paramétrique de la volatilité, et à réestimer le modèle ainsi contraint.

4.4.2 Un exemple (voir Gouriéroux et Monfort [1992])

Nous utilisons cette approche du général au spécifique pour l'étude des variations relatives journalières de l'indice de marché de la bourse de Paris (CAC40).

La première étape de spécification de la moyenne conditionnelle montre que celle-ci est approximativement nulle, ce qui est compatible avec l'hypothèse de marche aléatoire pour la valeur de l'indice en logarithme. Nous allons donc essentiellement considérer la spécification de la dynamique de la volatilité.

i) Asymétrie de la volatilité (« leverage effect »)

Nous commençons par chercher une spécification avec deux retards du type :

$$Y_t = h(Y_{t-1}, Y_{t-2}) u_t,$$

où u_t est un bruit blanc conditionnellement réduit, et approchons h par l'intermédiaire de fonctions étagées :

$$Y_t \simeq \left[\sum_{i=1}^6 \sum_{j=1}^6 \beta_{ij} \mathbf{1}_{A_i}(Y_{t-1}) \mathbf{1}_{A_j}(Y_{t-2}) \right] u_t,$$

où les ensembles A_1, \dots, A_6 correspondent aux intervalles de la partition définie par les valeurs limites :

$$-0.8\%, -0.4\%, 0\%, 0.4\%, 0.8\%.$$

Les valeurs estimées $\hat{\beta}_{ij}$ $i, j = 1, \dots, 6$ sont décrites ci-dessous.

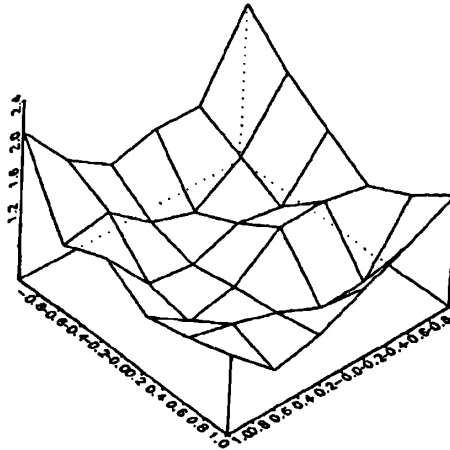


Figure 4.4: Volatilité estimée

Cette estimation semble montrer que la forme n'est pas exactement parabolique comme ce serait le cas avec une spécification ARCH usuelle, et qu'il y a une certaine asymétrie des réactions de la volatilité selon que la période correspond à une croissance de l'indice (Y_{t-1} et Y_{t-2} grands) ou à une décroissance de celui-ci. Ce point est confirmé par le tracé de la fonction donnant les valeurs $\hat{\beta}_{ii}$ (figure 4.5).

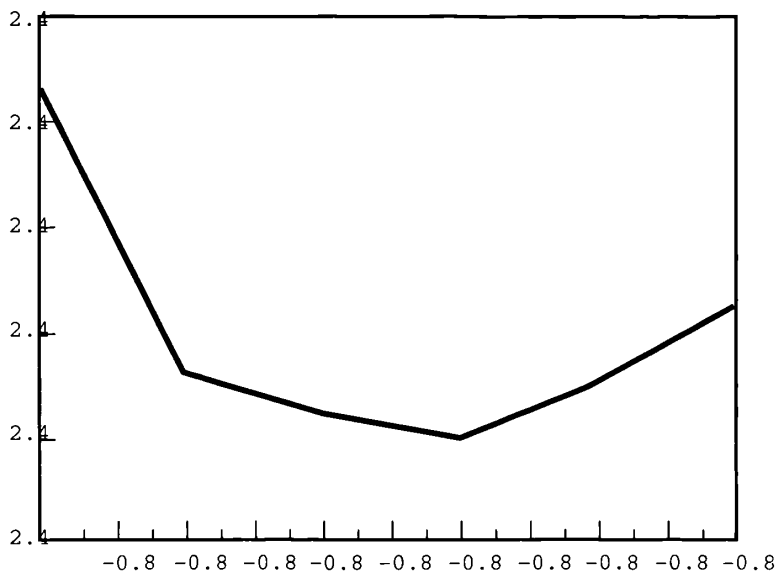


Figure 4.5: Volatilités estimées $\hat{\beta}_{ii}$

ii) Recherche d'une spécification contrainte

Il nous faut maintenant voir si cette asymétrie demeure, lorsqu'on augmente le nombre de retards. De façon à éviter une inflation de paramètres, nous retenons une formulation additive :

$$h_t = b_0 + \sum_{i=1}^3 \sum_{j=1}^4 b_{ij} \mathbf{1}_{A_i}(Y_{t-j}),$$

où les ensembles A_1, \dots, A_3 correspondent aux premiers intervalles de la partition limitée par :

$$-0.5\%, 0\%, 0.5\%.$$

Cette formulation est intuitivement introduite comme approximation de :

$$\begin{aligned} h_t &= h(Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}) \\ &= h_1(Y_{t-1}) + h_2(Y_{t-2}) + h_3(Y_{t-3}) + h_4(Y_{t-4}). \end{aligned}$$

Les coefficients b_{ij} estimés sont donnés ci-dessous retard par retard (figure 4.6).

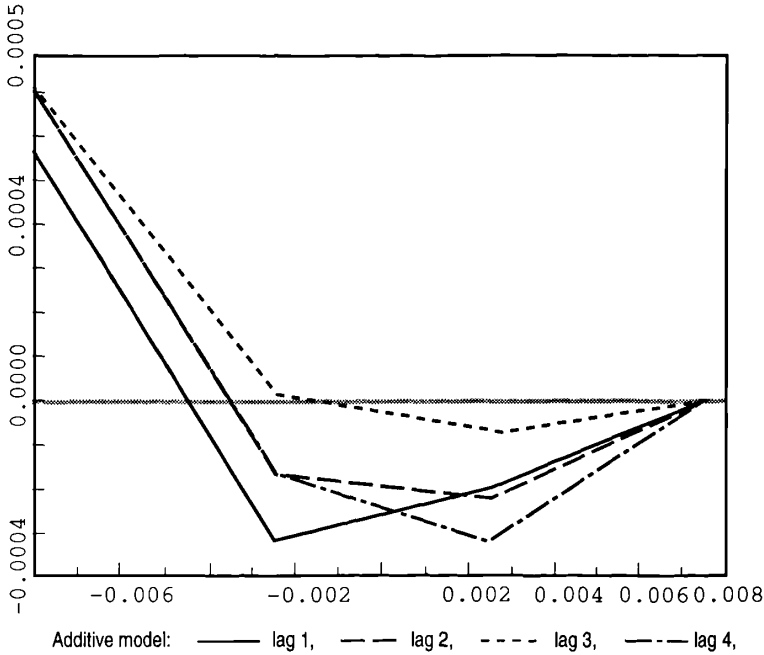


Figure 4.6: Coefficients estimés $\hat{\beta}_{ij}$

La figure montre un certain parallélisme des diverses courbes. Ceci incite à tester l'existence de liaisons entre les paramètres associés aux divers retards. Le test de l'hypothèse nulle :

$$H_0 = [\exists \lambda : b_{ij} = \lambda b_{i,j-1} \quad i = 1, 2, 3, j = 2, 3, 4]$$

conduit ainsi à accepter cette hypothèse.

On peut alors penser que :

$$\begin{aligned} h_t &\simeq h_1(Y_{t-1}) + \lambda h_1(Y_{t-2}) + \lambda^2 h_1(Y_{t-3}) + \lambda^3 h_1(Y_{t-4}) \\ &\simeq h_1(Y_{t-1}) + \lambda h_1(Y_{t-2}) + \lambda^2 h_1(Y_{t-3}) + \dots + \lambda^p h_1(Y_{t-p}) + \dots \end{aligned}$$

c'est-à-dire que :

$$h_t = \lambda h_{t-1} + h_1(Y_{t-1}).$$

La démarche conduit donc à une formulation GARCH(1,1) linéaire dans le terme h_{t-1} et non linéaire dans le terme Y_{t-1} du fait du phénomène d'asymétrie.

iii) Estimation de la spécification partiellement contrainte

L'estimation de la formulation :

$$h_t = \lambda h_{t-1} + h_1(Y_{t-1}),$$

par une approche de type noyau pour h_1 conduit à une fonction h_1 correspondant à la courbe en trait plein de la figure 4.7.

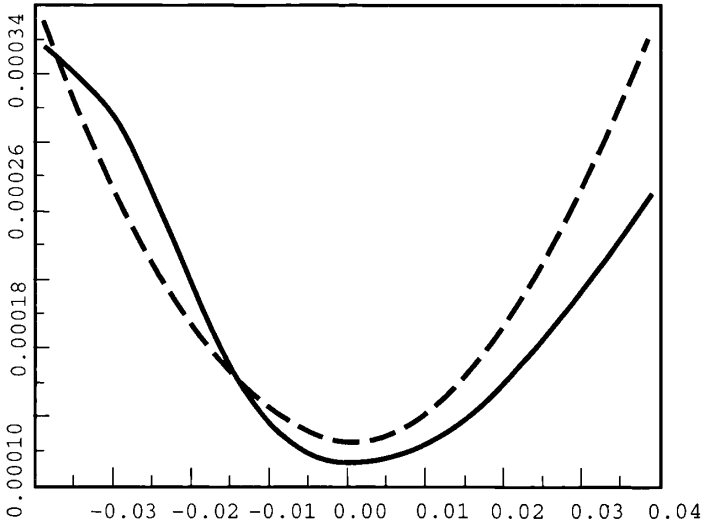


Figure 4.7: Influence de Y_{t-1} sur la volatilité

iv) Estimation d'une formulation GARCH(1,1)

L'estimation directe sur les données d'une spécification GARCH(1,1) conduit à :

$$h_t = 0.792 h_{t-1} + 0.159 Y_{t-1}^2 + 7.7 \cdot 10^{-6},$$

(13.2) (3.0) (2.3)

et à une fonction d'influence de Y_{t-1} sur h_t , de forme parabolique. La comparaison de cette courbe avec celle obtenue précédemment permet de préciser l'effet d'asymétrie et montre que l'utilisation a priori d'une dynamique quadratique risque d'entraîner des biais.

CHAPITRE 5

GÉNÉRALISATIONS ET ALTERNATIVES DE LA MODÉLISATION ARCH

par

Alain MONFORT

5.1 INTRODUCTION

Dans ce chapitre, on va d'abord examiner comment la notion de modèle ARCH ou GARCH univarié peut être généralisée au cas multivarié. On verra que, pour résoudre les problèmes posés par le grand nombre de paramètres, on est amené à introduire une notion générale de modèles dynamiques à facteurs, qui contient comme cas particulier la notion de modèle GARCH à facteurs endogènes et celle de modèle GARCH à facteurs exogènes. On examinera également certaines méthodes d'estimation adaptées à ces modèles ; des méthodes plus sophistiquées seront proposées dans le chapitre suivant. On étudiera ensuite les modèles à facteurs discrets et les modèles à volatilité stochastique discrète, ainsi que les méthodes d'inférence associées. Enfin, on s'intéressera aux modèles à volatilité stochastique continue.

5.2 PROCESSUS ARCH ET GARCH MULTIVARIÉS

5.2.1 Définitions générales

On a vu qu'un processus $GARCH(p, q)$ univarié était défini par les conditions :

$$\left\{ \begin{array}{l} E(\varepsilon_t/\varepsilon_{t-1}) = 0 \\ h_t = V(\varepsilon_t/\varepsilon_{t-1}) \\ = c + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j} \end{array} \right. \quad (5.1)$$

Lorsque le processus ε_t devient multivarié de taille n , il est naturel de proposer une généralisation directe des conditions précédentes :

$$\left\{ \begin{array}{l} E(\varepsilon_t/\varepsilon_{t-1}) = 0 \\ \text{vech}(H_t) = \text{vech} C + \sum_{i=1}^q A_i \text{vech}(\varepsilon_{t-i} \varepsilon_{t-i}') \\ \quad + \sum_{j=1}^p B_j \text{vech}(H_{t-j}) \\ \text{avec } H_t = V(\varepsilon_t/\varepsilon_{t-1}) \end{array} \right. \quad (5.2)$$

où vech désigne l'opérateur consistant à empiler les portions de colonnes d'une matrice symétrique se trouvant sur ou sous la diagonale principale, et où les A_i , B_j sont des matrices carrées de taille $\frac{n(n+1)}{2} \times \frac{n(n+1)}{2}$, tandis que C est une matrice symétrique $n \times n$.

On en déduit que le nombre de paramètres indépendants intervenant dans la définition précédente est :

$$\frac{n(n+1)}{2} + (p+q) \left[\frac{n(n+1)}{2} \right]^2.$$

Ainsi, dans le cas du modèle multivarié ARCH(1), c'est-à-dire dans le cas $p=0$, $q=1$, on obtient pour n variant de 1 à 4 : 2, 12, 42, 110.

Sur le plan de la stationarité on peut montrer (voir Bollerslev et Engle [1989], Gouriéroux [1992]) qu'une condition suffisante est que les racines du déterminant de $I - \sum_{i=1}^q A_i z^i - \sum_{j=1}^p B_j z^j$ soient de module supérieur à 1.

On conçoit que cette formulation générale pose rapidement un problème dû au trop grand nombre de paramètres. Par ailleurs la matrice H_t ainsi paramétrée n'est pas automatiquement semi-définie positive. Pour essayer de résoudre ces difficultés, divers modèles contraints ont été proposés.

Les modèles *diagonaux* (Bollerslev, Engle et Wooldridge [1988]) sont des modèles dans lesquels les matrices A_j , B_j sont diagonales. Cette formulation est

évidemment très simple mais elle présente divers inconvénients : absence de stabilité par transformation de ε_t par une application linéaire et faible compatibilité avec la condition de positivité (voir Gouriéroux [1992]).

Les modèles à *corrélations conditionnelles constantes* (Bollerslev [1987]) sont définis par :

$$\begin{cases} h_{kkt} = c_k + \sum_{i=1}^q a_{ik} \varepsilon_{k,t-i}^2 + \sum_{j=1}^p b_{jk} h_{kk,t-j} \\ h_{k\ell t} = \rho_{k\ell} h_{kkt}^{1/2} h_{\ell\ell t}^{1/2} \quad k \neq \ell \end{cases} \quad (5.3)$$

$$H_t = \text{diag} \left(h_{kkt}^{1/2} \right) R \text{diag} \left(h_{kkt}^{1/2} \right)$$

R étant une matrice de corrélation ou encore :

$$H_t = \text{diag} \left(d_{kkt}^{1/2} \right) \Gamma \text{diag} \left(d_{kkt}^{1/2} \right)$$

où $d_{kkt} = h_{kkt}/c_k$ et Γ est une matrice de variance-covariance.

Il est clair que la contrainte de positivité est satisfaite dès que les variances h_{kkt} sont positives. Cependant la stabilité par transformation linéaire n'est pas satisfaite.

Enfin, Baba, Engle, Kraft et Kroner [1989] ont proposé une formulation du type :

$$H_t = C + \sum_{k=1}^K \sum_{i=1}^q A_{ik}^* \varepsilon_{t-i} \varepsilon'_{t-i} A_{ik}^* + \sum_{k=1}^K \sum_{j=1}^p B_{ik}^* H_{t-j} B_{ik}^* \quad (5.4)$$

où les matrices A_{ik}^* , B_{jk}^* sont des matrices carrées $n \times n$. Cette formulation assure que H_t est semi-définie positive (si C est semi-définie positive) et elle est stable par transformation linéaire. Cependant une telle formulation peut encore nécessiter un grand nombre de paramètres et son interprétation n'est pas évidente.

Pour aller plus loin, il faut introduire les modèles ARCH à facteurs et, pour cela, nous allons d'abord nous intéresser à la notion générale de modèles dynamiques à facteurs.

5.2.2 Modèles dynamiques à facteurs

On dira qu'un processus multivarié Y_t de taille n a une représentation factorielle d'ordre r , ou suit un modèle à facteurs d'ordre r , si on a :

$$Y_t = \beta \alpha' X_t + u_t \quad (5.5)$$

où $E(u_t/J_{t-1}) = 0$, $\text{cov}(u_t, X_t/J_{t-1}) = E(u_t X_t'/J_{t-1}) = 0$, $V(u_t/J_{t-1}) = \Omega$,

$$J_{t-1} = \{Y_{t-1}, X_{t-1}, Y_{t-2}, X_{t-2}, \dots\},$$

où X_t est un vecteur de taille K et où les matrices β , α sont de tailles $(n \times r)$, $(K \times r)$ et de rang r .

Cette définition signifie que la dynamique du vecteur Y_t est, pour les deux premiers moments, contenue dans celle du processus $F_t = \alpha \cdot X_t$, dont les composantes sont appelés facteurs exogènes, puisque :

$$E(Y_t/J_{t-1}) = \beta E(F_t/J_{t-1})$$

et

$$V(Y_t/J_{t-1}) = \beta V(F_t/J_{t-1}) \beta' + \Omega. \quad (5.6)$$

On vérifie d'ailleurs que si le processus Y_t a une représentation factorielle d'ordre r pour l'information $J_{t-1} = \{Y_{t-1}, X_{t-1}, Y_{t-2}, X_{t-2}, \dots\}$, il admet la même représentation factorielle d'ordre r pour toute information I_{t-1} incluse dans J_{t-1} . En effet, on a :

$$E(u_t/I_{t-1}) = E[E(u_t/J_{t-1})/I_{t-1}] = 0$$

$$\begin{aligned} \text{Cov}(u_t, X_t/I_{t-1}) &= E(u_t X_t'/I_{t-1}) \\ &= E[E(u_t X_t'/J_{t-1})/I_{t-1}] = 0 \end{aligned}$$

$$V(u_t/I_{t-1}) = E[V(u_t/J_{t-1})/I_{t-1}] + V[E(u_t/J_{t-1})/I_{t-1}] = \Omega.$$

Cette propriété est donc vraie en particulier pour $I_{t-1} = \{Y_{t-1}, Y_{t-2}, \dots\} = \underline{Y}_{t-1}$ et ceci implique donc :

$$\begin{cases} E(Y_t/\underline{Y}_{t-1}) = \beta E(F_t/\underline{Y}_{t-1}) \\ V(Y_t/\underline{Y}_{t-1}) = \beta V(F_t/\underline{Y}_{t-1}) \beta' + \Omega. \end{cases} \quad (5.7)$$

On peut aussi montrer (voir Gourieroux, Monfort et Renault [1991b]) que le processus Y_t a une représentation factorielle d'ordre r , si et seulement si il a une représentation factorielle d'ordre r endogène, c'est-à-dire du type :

$$Y_t = \beta F_t^* + u_t^*$$

avec :

$$\begin{aligned} E(u_t^*/J_{t-1}) &= 0, \\ \text{Cov}(u_t^*, F_t^*/J_{t-1}) &= 0, \\ V(u_t^*/J_{t-1}) &= \Omega^* (= \Omega - \beta(\beta' \Omega^{-1} \beta)^{-1} \beta'), \\ F_t^* &= a' Y_t, \end{aligned}$$

a étant une matrice de taille $(n \times r)$ de rang r définie par $a' = (\beta' \Omega^{-1} \beta)^{-1} \beta' \Omega^{-1}$. Les composantes de F_t^* sont appelés facteurs endogènes. Notons que $a' u_t^* = 0$

et donc que le noyau de Ω^* est engendré par les colonnes de a . On voit également que le choix de a et β n'est pas unique puisqu'on peut toujours remplacer β par βQ et a' par $Q^{-1}a'$; on peut imposer, par exemple, que la matrice ($r \times r$) supérieure de β soit l'identité.

Les formules (5.6) impliquent donc :

$$\begin{aligned} E(Y_t / J_{t-1}) &= \beta E(F_t^* / J_{t-1}) \\ V(Y_t / J_{t-1}) &= \beta V(F_t^* / J_{t-1})\beta' + \Omega^* \end{aligned} \quad (5.6')$$

et les formules (5.7) entraînent aussi :

$$\begin{aligned} E\left(Y_t / Y_{t-1}\right) &= \beta E\left(F_t^* / Y_{t-1}\right) \\ V\left(Y_t / Y_{t-1}\right) &= \beta V\left(F_t^* / Y_{t-1}\right)\beta' + \Omega^*. \end{aligned} \quad (5.7')$$

Revenons maintenant au cas des paragraphes précédents où on suppose en plus que $E\left(Y_t / Y_{t-1}\right) = 0$ et reprenons la notation ε_t au lieu de Y_t . Les formules (5.7) ou (5.7)' impliquent que $H_t = V\left(\varepsilon_t / \varepsilon_{t-1}\right)$ s'écrit sous la forme $\Omega + \beta V\left(F_t^* / \varepsilon_{t-1}\right)\beta'$ ou $\Omega^* + \beta V\left(F_t^* / \varepsilon_{t-1}\right)\beta'$. Pour spécifier davantage H_t , il faut maintenant faire des hypothèses soit sur le processus des facteurs endogènes, soit sur le processus des facteurs exogènes.

5.2.3 Modèles GARCH à facteurs endogènes

Comme la notion de modèle à facteurs endogènes n'impose pas de restriction sur la dynamique des facteurs, on peut supposer que les r facteurs endogènes, c'est-à-dire les r composantes de $F_t^* = a'\varepsilon_t$, sont des processus GARCH(p, q) indépendants dont les variances conditionnelles sont données par :

$$v_{k,t} = c_k + \sum_{i=1}^q \lambda_{ik}^2 (a'_k \varepsilon_{t-i})^2 + \sum_{j=1}^p \phi_{jk}^2 v_{k,t-j}. \quad (5.8)$$

La matrice $V\left(F_t^* / \varepsilon_{t-1}\right)$ est donc la matrice $\text{diag}(v_{k,t})$ et la formule (5.7)' fournit l'expression suivante pour H_t :

$$H_t = \Omega^* + \beta \text{diag}(c_k)\beta' + \sum_{k=1}^r \sum_{i=1}^q \lambda_{ik}^2 \beta_k \beta'_k (a'_k \varepsilon_{t-i})^2 + \sum_{k=1}^r \sum_{j=1}^p \phi_{jk}^2 \beta_k \beta'_k v_{k,t-j}$$

Comme $v_{k,t-j} = V \left(a'_k \varepsilon_{t-j} / \varepsilon_{t-j-1} \right) = a'_k H_{t-j} a_k$, on a :

$$H_t = \tilde{\Omega} + \sum_{k=1}^r \sum_{i=1}^q \lambda_{ik}^2 \beta_k \beta'_k (a'_k \varepsilon_{t-i})^2 + \sum_{k=1}^r \sum_{j=1}^p \phi_{jk}^2 \beta_k \beta'_k a'_k H_{t-j} a_k$$

avec :

$$\tilde{\Omega} = \Omega^* + \beta \text{diag}(c_k) \beta'$$

ou encore :

$$H_t = \tilde{\Omega} + \sum_{k=1}^r \sum_{i=1}^q \lambda_{ik}^2 \beta_k a'_k \varepsilon_{t-i} \varepsilon'_{t-i} a_k \beta'_k + \sum_{k=1}^r \sum_{j=1}^p \phi_{jk}^2 \beta_k a'_k H_{t-j} a_k \beta'_k. \quad (5.9)$$

On obtient donc une forme du type (5.4) avec des matrices A_{ik}^* et B_{jk}^* de rang 1 données par :

$$A_{ik}^* = \lambda_{ik} a_k \beta'_k, \quad B_{jk}^* = \phi_{jk} a_k \beta'_k.$$

Notons que l'expression $a' = (\beta' \Omega^{-1} \beta)^{-1} \beta' \Omega^{-1}$ implique les contraintes $a' \beta = I$, c'est-à-dire $a'_k \beta_\ell = 0$, $\forall k \neq \ell$, et $a'_k \beta_k = 1$.

Une formulation de ce type a été proposée par Engle [1987], avec une normalisation différente du vecteur des facteurs fournissant des corrélations fixes mais non nulles. Les contraintes de non corrélation des facteurs s'écrivent $a'_k \tilde{\Omega} a_\ell = 0$, $\forall k \neq \ell$ (ce qui est vérifié si on impose que le noyau de Ω^* soit engendré par les a_k , $k = 1, \dots, r$). Rappelons enfin que des contraintes d'identification supplémentaires sont nécessaires, par exemple que la matrice $(r \times r)$ supérieure de β soit l'identité.

5.2.4 Modèles GARCH à facteurs exogènes

Ce type de modèle a été proposé par Diebold-Nerlove [1989]. Considérons, pour simplifier, le modèle à un facteur ; ce modèle s'écrit :

$$\varepsilon_t = \lambda F_t + e_t$$

où $\{e_t\}$ est un bruit blanc gaussien de matrice variance-covariance Γ , que Diebold-Nerlove supposent diagonale, où $\{F_t\}$ est un processus scalaire GARCH (p, q) (de variance marginale égale à 1 de façon à rendre les paramètres identifiables), indépendant du processus $\{e_t\}$, et où λ est un vecteur de paramètres inconnus.

On vérifie facilement que ce modèle est un modèle à facteurs exogènes dont la matrice de variance-covariance conditionnelle est, par exemple dans le cas où F_t est un ARCH(1) de variance conditionnelle $1 - \alpha + \alpha F_{t-1}^2$ ($0 < \alpha < 1$) :

$$\begin{aligned}
 H_t &= \mathbf{V}(\varepsilon_t / \varepsilon_{t-1}) \\
 &= \mathbf{V} \left[\mathbf{E}(\varepsilon_t / F_{t-1}, e_{t-1}) / \varepsilon_{t-1} \right] + \mathbf{E} \left[\mathbf{V}(\varepsilon_t / F_{t-1}, e_{t-1}) / \varepsilon_{t-1} \right] \\
 &= \Gamma + \lambda \lambda' \mathbf{E} \left[\mathbf{V}(F_t / F_{t-1}, e_{t-1}) / \varepsilon_{t-1} \right] \\
 &= \Gamma + \lambda \lambda' \left[1 - \alpha + \alpha \mathbf{E}(F_{t-1}^2 / \varepsilon_{t-1}) \right].
 \end{aligned}$$

Cette matrice a bien la forme générale (5.7) puisque $\mathbf{V}(F_t / \varepsilon_{t-1}) = 1 - \alpha + \alpha \mathbf{E}(F_{t-1}^2 / \varepsilon_{t-1})$, mais elle n'a pas d'expression explicite en fonction de ε_{t-1} .

5.2.5 Estimation des modèles GARCH à facteurs

On a vu que, pour les modèles GARCH à facteurs endogènes, on dispose d'une expression explicite de $H_t(\varepsilon_t / \varepsilon_{t-1})$. Par conséquent, on peut calculer la (pseudo) log-vraisemblance du modèle :

$$-\frac{Tn}{2} \text{Log } 2\pi - \frac{1}{2} \sum_{t=1}^T (\text{Log } \det H_t + \varepsilon_t' H_t^{-1} \varepsilon_t).$$

On peut donc, en principe, maximiser cette (pseudo) log-vraisemblance même si le choix du nombre de facteurs et la prise en compte des contraintes d'identification peut être délicate (voir Lin [1991]). Il faut aussi noter qu'en pratique, des considérations économiques peuvent conduire à remplacer les facteurs endogènes inobservables $a'_k \varepsilon_t$ par des observations (par exemple un « portefeuille de marché ») (voir Engle, Ng et Rothschild [1989], [1990], King, Sentana et Wadhvani [1990]).

L'estimation des modèles GARCH à facteurs exogènes est plus complexe puisqu'on ne dispose pas d'une forme explicite pour H_t . On va décrire ici deux méthodes approchées qui, utilisées seules, conduisent à des estimateurs dont les propriétés asymptotiques ne sont pas claires mais qui, utilisées dans le cadre de l'estimation indirecte (voir le chapitre 6), fournissent des estimateurs convergents et asymptotiquement normaux.

La première méthode approchée est fondée sur le filtre de Kalman étendu. En effet le modèle peut s'écrire sous la forme espace d'états non linéaire courante (dans le cas où F est un ARCH(1)) :

$$\begin{cases} F_t = [1 + \alpha (F_{t-1}^2 - 1)]^{1/2} \eta_t & (\text{équation de transition}) \\ \varepsilon_t = \lambda F_t + e_t & (\text{équation de mesure}) \end{cases}$$

où $\{\eta_t\}$ est un bruit blanc gaussien centré de variance 1 et $\{e_t\}$ un bruit blanc gaussien de matrice de variance-covariance Γ , indépendant de $\{\eta_t\}$.

Le filtre de Kalman étendu conduit à un calcul récursif d'une approximation \tilde{H}_t de H_t , fondé sur les équations :

$$\begin{cases} F_{t/t} = (1 - \alpha + \alpha F_{t-1/t-1}^2) \lambda' \tilde{H}_t^{-1} \varepsilon_t \\ \tilde{H}_t = (1 - \alpha + \alpha F_{t-1/t-1}^2) \lambda \lambda' + \Gamma. \end{cases}$$

On peut alors calculer une approximation de la log-vraisemblance.

La deuxième méthode approchée repose sur une discrétisation de l'espace des valeurs de F_t . Plus précisément on partitionne \mathbb{R} en K intervalles :

$[a_k, a_{k+1}[$, $k = 0, \dots, K - 1$, $a_0 = -\infty$, $a_K = +\infty$ et on pose :

$$\tilde{F}_t = b_k \text{ si } F_t \in [a_k, a_{k+1}[$$

b_k étant une valeur de $[a_k, a_{k+1}[$, par exemple le milieu pour les intervalles bornés.

On a alors :

$$\begin{aligned} P[\tilde{F}_t = b_k / \tilde{F}_{t-1} = b_\ell] &= P[F_t \in [a_k, a_{k+1}[/ F_{t-1} \in [a_\ell, a_{\ell+1}[] \\ &\simeq P(F_t \in [a_k, a_{k+1}[/ F_{t-1} = b_\ell) \\ &= \Phi \left[\frac{a_{k+1}}{(1 - \alpha + \alpha b_\ell^2)^{1/2}} \right] - \Phi \left[\frac{a_k}{(1 - \alpha + \alpha b_\ell^2)^{1/2}} \right] \end{aligned}$$

où Φ est la fonction de répartition de $N(0, 1)$.

F_t est donc approché par la chaîne de Markov \tilde{F}_t dont les probabilités de transition sont fonctions des paramètres d'intérêt. On peut alors calculer la log-vraisemblance exacte du modèle approché $\varepsilon_t = \lambda \tilde{F}_t + e_t$ par des techniques de filtrage non linéaire proposées par Hamilton [1989] (voir la section suivante).

5.3 MODÈLES À FACTEURS DISCRETS ET MODÈLES À VOLATILITÉ STOCHASTIQUE DISCRÈTE

5.3.1 Définition

Soit S_t une chaîne de Markov homogène à K états, notés $1, 2, \dots, K$, dont les probabilités de transitions sont notées $p_{k\ell}$ $k = 1, \dots, K$, $\ell = 1, \dots, K$. Hamilton [1988, 1989] a étudié des modèles du type :

$$Y_t - \mu(S_t) = \phi_1[Y_{t-1} - \mu(S_{t-1})] + \dots + \phi_m[Y_{t-m} - \mu(S_{t-m})] + \sigma(S_t)v_t$$

où $\{v_t\}$ suit IIN(0, 1) indépendamment de $\{S_t\}$ et où $\mu(S_t)$ et $\sigma(S_t)$ sont des fonctions définies sur $\{1, 2, \dots, K\}$ à valeurs dans \mathbb{R} et \mathbb{R}^+ respectivement.

Si on pose

$$Z_t = Y_t - \mu(S_t)$$

on a :

$$\text{Phi}(L)Z_t = \sigma(S_t)v_t$$

où $\Phi(L) = 1 - \phi_1 L - \dots - \phi_m L^m$ et L désigne l'opérateur retard. Donc le modèle peut s'écrire :

$$Y_t = \mu(S_t) + \Phi^{-1}(L)\sigma(S_t)v_t.$$

Conditionnellement au processus $\{S_t\}$, le processus $\{Y_t\}$ est donc un $AR(m)$ avec un terme constant et une variance de l'innovation dépendant de t . Pour adapter ce modèle à nos besoins, nous allons à la fois le particulariser en prenant $\Phi(L) = I$ et le généraliser en rendant Y_t multivarié (de taille n). On pourra aussi imposer que Y_t soit une différence de martingale et on reprend la notation ε_t au lieu de Y_t . Finalement le modèle auquel on s'intéressera est :

$$\varepsilon_t = \lambda\mu(S_t) + \Sigma^{1/2}(S_t)v_t \tag{5.11}$$

où λ est un vecteur de taille n , $\Sigma^{1/2}(S_t)$ une application de $\{1, \dots, K\}$ dans l'ensemble des matrices $(n \times n)$ symétriques définies positives et v_t suit IIN(0, I_d) indépendamment de S_t .

Ce modèle contient comme cas particuliers les deux modèles d'intérêt : d'une part le modèle à un facteur discret si $\Sigma^{1/2}(S_t)$ est fixe et notée $\Sigma^{1/2}$:

$$\varepsilon_t = \lambda\mu(S_t) + \Sigma^{1/2}v_t \tag{5.12}$$

et les modèles univariés à volatilité stochastique discrète, obtenu en faisant $n = 1$ et $\lambda = 0$:

$$\varepsilon_t = \sigma(S_t)v_t. \tag{5.13}$$

Notons que, dans le modèle à volatilité stochastique discrète, ε_t est automatiquement une différence de martingale puisque :

$$E(\varepsilon_t / \varepsilon_{t-1}) = E[E(\varepsilon_t / S_t, v_{t-1}) / \varepsilon_{t-1}] = 0.$$

Dans le premier modèle cette condition sera vérifiée si $\mu(S_t)$ est une différence de martingale, ce qui implique les contraintes suivantes sur les paramètres $\mu_k, p_{k\ell}$ (μ_k étant une notation simplifiée de $\mu(k)$) :

$$\sum_{k=1}^K p_{k\ell} \mu_k = 0 \quad \ell = 1, \dots, K.$$

Notons que dans le modèle à facteurs il faut imposer une contrainte d'identification du terme $\lambda\mu(S_t)$, par exemple $\mu_1 = 1$.

5.3.2 Filtrage, lissage, estimation

i) Algorithme de type Kalman

On va proposer un algorithme récursif analogue au filtre de Kalman, dont l'entrée est la probabilité conditionnelle :

$$p\left(S_{t-1} = k/\varepsilon_{t-1}\right), \quad k = 1, \dots, K$$

et dont la sortie est la même probabilité pour la date t , $p\left(S_t = k/\varepsilon_t\right)$; cet algorithme fournit également la densité $p\left(\varepsilon_t/\varepsilon_{t-1}\right)$.

L'algorithme se décompose en quatre étapes (avec des notations évidentes) :

1ère étape :

$$\begin{aligned} p\left(s_t, s_{t-1}/\varepsilon_{t-1}\right) &= p\left(s_t/s_{t-1}, \varepsilon_{t-1}\right) p\left(s_{t-1}/\varepsilon_{t-1}\right) \\ &= p\left(s_t/s_{t-1}\right) p\left(s_{t-1}/\varepsilon_{t-1}\right) \\ p\left(s_t/\varepsilon_{t-1}\right) &= \sum_{s_{t-1}=1}^K p\left(s_t/s_{t-1}\right) p\left(s_{t-1}/\varepsilon_{t-1}\right). \end{aligned}$$

2ème étape :

$$p\left(\varepsilon_t, s_t/\varepsilon_{t-1}\right) = p\left(\varepsilon_t/s_t, \varepsilon_{t-1}\right) p\left(s_t/\varepsilon_{t-1}\right)$$

le premier terme du membre de droite étant :

$$(2\pi)^{-\frac{n}{2}} \det \Sigma^{-1/2}(s_t) \exp \left[-\frac{1}{2} (\varepsilon_t - \lambda\mu(s_t))' \Sigma^{-1}(s_t) (\varepsilon_t - \lambda\mu(s_t)) \right].$$

3ème étape :

$$p\left(\varepsilon_t/\varepsilon_{t-1}\right) = \sum_{s_t=1}^K p\left(\varepsilon_t, s_t/\varepsilon_{t-1}\right).$$

4ème étape :

$$p\left(s_t/\varepsilon_t\right) = \frac{p\left(\varepsilon_t, s_t/\varepsilon_{t-1}\right)}{p\left(\varepsilon_t/\varepsilon_{t-1}\right)}.$$

L'algorithme fournit les probabilités de prévision de S_t $p\left(s_t/\varepsilon_{t-1}\right)$ et de filtrage de S_t $p\left(s_t/\varepsilon_t\right)$ ainsi que les probabilités de prévision des observations

$p(\varepsilon_t/\varepsilon_{t-1})$. Ces dernières permettent en particulier de calculer la log-vraisemblance pour des valeurs données des paramètres.

L'initialisation de l'algorithme nécessiterait le calcul de $p(s_1/y_1)$ qui est remplacé, en pratique, par la probabilité invariante de la chaîne de Markov.

On peut aussi utiliser la technique de Kitagawa [1987] pour calculer les probabilités de lissage $p(s_t/\varepsilon_T)$. En effet on a :

$$\begin{aligned} p(s_t, s_{t+1}/\varepsilon_T) &= p(s_{t+1}/\varepsilon_T) p(s_t/s_{t+1}, \varepsilon_T) \\ &= p(s_{t+1}/\varepsilon_T) p(s_t/s_{t+1}, \varepsilon_t) \\ &= p(s_{t+1}/\varepsilon_T) \frac{p(s_t, s_{t+1}/\varepsilon_t)}{p(s_{t+1}/\varepsilon_t)} \\ &= \frac{p(s_{t+1}/\varepsilon_T) p(s_{t+1}/s_t) p(s_t/\varepsilon_t)}{p(s_{t+1}/\varepsilon_t)}. \end{aligned} \tag{5.14}$$

Par conséquent :

$$p(s_t/\varepsilon_T) = \sum_{s_{t+1}=1}^K \frac{p(s_{t+1}/\varepsilon_T) p(s_{t+1}/s_t) p(s_t/\varepsilon_t)}{p(s_{t+1}/\varepsilon_t)}.$$

Ce qui fournit un algorithme récursif descendant pour les $p(s_t/\varepsilon_T)$, pourvu que les probabilités de prévision et de filtrage aient été stockées.

Ces probabilités de lissage contiennent donc l'information optimale sur la trajectoire de la chaîne ; elles permettent en particulier de calculer les espérances conditionnelles $E(s_t/\varepsilon_T)$. Dans le cas du modèle à facteurs, on peut donc obtenir ainsi la trajectoire lissée du facteur inobservable.

Notons enfin, que, lorsque les probabilités de lissage $p(s_{t+1}/\varepsilon_T)$ ont été calculées, on déduit immédiatement de (5.14) les probabilités de lissage jointes $p(s_t, s_{t+1}/\varepsilon_T)$.

ii) Algorithme EM

Une autre approche consiste à utiliser l'algorithme EM (espérance maximisation) (Dempster-Laird-Rubin[1977]). Cet algorithme est le suivant. Soit $\theta^{(j)}$, la valeur du paramètre lors de l'itération j , la valeur du paramètre à l'itération $j + 1$, $\theta^{(j+1)}$ est donnée par la maximisation en θ de :

$$E_{\theta^{(j)}}[L_T^*(\varepsilon_T, s_T, \theta)/\varepsilon_T]$$

où $L_T^*(\varepsilon_T, s_T, \theta)$ est la log-densité de (ε_T, s_T) .

Dans le cas particulier qui nous intéresse, l'algorithme EM a la propriété remarquable de fournir une forme explicite pour $\theta^{(j+1)}$ en fonction des probabilités de lissage.

Ainsi, dans le cas du modèle à un facteur on a :

$$L_T^* = -\frac{Tn}{2} \text{Log } 2\pi - \frac{T}{2} \text{Log det } \Sigma$$

$$-\frac{1}{2} \sum_{t=1}^T (\varepsilon_t - \lambda\mu(s_t))' \Sigma^{-1} (\varepsilon_t - \lambda\mu(s_t)) + \sum_{t=1}^T \text{Log } p(s_t/s_{t-1})$$

donc :

$$\frac{\partial L_T^*}{\partial \lambda} = \sum_{t=1}^T \Sigma^{-1} (\mu(s_t)\varepsilon_t - \lambda\mu^2(s_t)).$$

En annulant l'espérance conditionnelle de $\frac{\partial L_T^*}{\partial \lambda}$ sachant ε_T et pour les valeurs des paramètres à l'itération j , on obtient :

$$\lambda = \frac{\sum_{t=1}^T \sum_{k=1}^K \varepsilon_t \mu_k p^{(j)}(S_t = k/\varepsilon_T)}{\sum_{t=1}^T \sum_{k=1}^K \mu_k^2 p^{(j)}(S_t = k/\varepsilon_T)}. \quad (5.15)$$

De même :

$$\frac{\partial L_T^*}{\partial \Sigma^{-1}} = \frac{T}{2} \Sigma - \frac{1}{2} \sum_{t=1}^T (\varepsilon_t - \lambda\mu(s_t))(\varepsilon_t - \lambda\mu(s_t))'$$

et en annulant l'espérance conditionnelle :

$$\Sigma = \frac{1}{T} \sum_{t=1}^T (\varepsilon_t - \lambda\mu_k)(\varepsilon_t - \lambda\mu_t)' p^{(j)}(S_t = k/\varepsilon_T). \quad (5.16)$$

On a aussi :

$$\frac{\partial L_T^*}{\partial \mu_k} = \sum_{t=1}^T \lambda' \Sigma^{-1} (\varepsilon_t - \lambda\mu_k) \mathbf{1}_{\{k\}}(s_t)$$

d'où en annulant l'espérance conditionnelle :

$$\mu_k = \frac{\lambda' \Sigma^{-1} \sum_{t=1}^T \varepsilon_t p^{(j)}(S_t = k/\varepsilon_T)}{\lambda' \sum_{t=1}^T \lambda \sum_{k=1}^K p^{(j)}(S_t = k/\varepsilon_T)} \quad k = 2, \dots, K. \quad (5.17)$$

Enfin, pour $k = 1, \dots, K - 1, \ell = 1, \dots, K$:

$$\frac{\partial L_T^*}{\partial p_{k\ell}} = \sum_{t=1}^T \left[\mathbf{1}_{\{k,\ell\}}(s_t, s_{t-1}) \frac{1}{p_{k\ell}} - \mathbf{1}_{\{K,\ell\}}(s_t, s_{t-1}) \frac{1}{p_{K\ell}} \right]$$

d'où en annulant l'espérance conditionnelle :

$$p_{k\ell} = \frac{\sum_{t=1}^T p^{(j)}(S_t = k, S_{t-1} = \ell / \varepsilon_T)}{\sum_{t=1}^T p^{(j)}(S_{t-1} = \ell / \varepsilon_T)} \quad k = 1, \dots, K, \ell = 1, \dots, K. \quad (5.18)$$

Les valeurs des $p_{k\ell}$ à l'itération $(j + 1)$ sont fournies directement par ces dernières équations ; les valeurs de λ, Σ et μ_k sont obtenues en résolvant les équations (5.15) (5.16) (5.17). Notons que lorsque les μ_k sont connus, ce qui est le cas lorsque le modèle est utilisé comme approximation d'un modèle à facteur à valeurs réelles (voir 5.2.5), on obtient les valeurs de λ et Σ à l'itération $(j + 1)$ de façon immédiate puisque les équations (5.15) (5.16) sont alors récursives. Il faut cependant remarquer que, dans le modèle obtenu en 5.2.5, les probabilités de transition $p_{k\ell}$ ne sont pas libres puisqu'elles sont fonctions d'un seul paramètre α . Pour continuer à bénéficier de la simplicité de calcul que l'on vient de mentionner, une solution pourrait donc être d'utiliser une méthode d'estimation indirecte (voir le chapitre 6) dont la fonction critère est obtenue à partir du modèle discrétisé non contraint.

5.4 MODÈLES À VOLATILITÉ STOCHASTIQUE CONTINUE

5.4.1 Cas univarié

Harvey, Ruiz et Shephard [1992] proposent le modèle suivant :

$$\begin{cases} Y_t = u_t \sigma_t = u_t \exp\left(\frac{1}{2} h_t\right) \\ h_t = \gamma + \phi h_{t-1} + v_t \end{cases} \quad (5.19)$$

où $\{u_t\}$ et $\{v_t\}$ sont des bruits blancs gaussiens indépendants de variances respectives 1 et σ_v^2 . Y_t représente par exemple le rendement d'un actif financier, l'accroissement relatif d'un taux de change...

Pour pouvoir traiter ce type de modèle par les méthodes récursives de Kalman standard, on peut mettre la première équation sous la forme :

$$Y_t^2 = u_t^2 \exp(h_t)$$

ou :

$$\text{Log } Y_t^2 = h_t + \text{Log } u_t^2.$$

L'espérance et la variance de $\text{Log } u_t^2$ valent respectivement -1.27 et 4.93 et si on pose $\xi = \text{Log } u_t^2 + 1.27$ on obtient un bruit blanc centré (non gaussien) de variance 4.93 . Au total on peut réécrire le modèle (5.19) sous la forme :

$$\begin{cases} \text{Log } Y_t^2 = -1.27 + h_t + \xi_t, \\ h_t = \gamma + \phi h_{t-1} + v_t. \end{cases} \quad (5.20)$$

On reconnaît une forme espace d'états linéaire. Le bruit ξ_t n'est pas gaussien, mais on peut utiliser le filtre de Kalman standard pour calculer la pseudo-vraisemblance (voir Gourieroux, Monfort et Trognon [1984]) associée au cas normal. La maximisation de cette pseudo-vraisemblance fournit des estimateurs convergents et asymptotiquement normaux ; cependant la matrice de variance-covariance asymptotique n'est pas celle fournie par la théorie du maximum de vraisemblance mais celle fournie par la théorie du pseudo-maximum vraisemblance à savoir $J_0 I_0^{-1} J_0$ avec

$$\begin{aligned} J_0 &= \lim_{T \rightarrow \infty} -\frac{1}{T} \frac{\partial^2 L_T}{\partial \theta \partial \theta'} \\ I_0 &= \lim_{T \rightarrow \infty} \frac{1}{T} V_0 \left(\frac{\partial L_T(\theta_0)}{\partial \theta} \right) \end{aligned} \quad (5.21)$$

où L_T est la pseudo-logvraisemblance, $\hat{\theta}_T$ l'estimateur du pseudo-maximum de vraisemblance et θ_0 la vraie valeur du paramètre. J_0 peut être estimée par

$$-\frac{1}{T} \frac{\partial^2 L_T}{\partial \theta \partial \theta'}(\hat{\theta}_T)$$

et I_0 par (voir Newey et West [1987]) :

$$\begin{aligned} & \frac{1}{T} \sum_{t=1}^T \frac{\partial \ell_t}{\partial \theta}(\hat{\theta}_T) \frac{\partial \ell_t}{\partial \theta'}(\hat{\theta}_T) + \frac{1}{T} \sum_{j=1}^{k_T} \left(1 - \frac{j}{k_T + 1} \right) \\ & \times \sum_{t=j+1}^T \left(\frac{\partial \ell_t}{\partial \theta}(\hat{\theta}_T) \frac{\partial \ell_{t-j}}{\partial \theta'}(\hat{\theta}_T) + \frac{\partial \ell_{t-j}}{\partial \theta}(\hat{\theta}_T) \frac{\partial \ell_t}{\partial \theta'}(\hat{\theta}_T) \right) \end{aligned}$$

où

$$L_T = \sum_{t=1}^T \ell_t$$

et k_T est une suite d'entiers tendant vers $+\infty$ = plus lentement que $T^{1/4}$.

5.4.2 Cas multivarié

Si n est la dimension de Y_t , une généralisation naturelle de (5.19) est :

$$\begin{cases} Y_{it} = u_{it} \exp(\frac{1}{2}h_{it}) \\ h_{it} = \gamma_i + \phi_1 h_{i,t-1} + v_{it} \quad i = 1, \dots, n \end{cases} \quad (5.22)$$

où $\{u_t\}$, vecteur de composantes u_{it} , est un bruit blanc gaussien de matrice de variance-covariance Σ_u dont les termes diagonaux valent 1 (pour que le modèle soit identifiable) et $\{v_t\}$, vecteur de composantes v_{it} , est un bruit blanc gaussien de matrice de variance-covariance Σ_v et indépendant du processus $\{u_t\}$.

En notant Z_t le vecteur dont les composantes sont $\text{Log } Y_{it}^2$ ($i = 1, \dots, n$) et ξ_t le vecteur dont les composantes sont $1.27 + \text{Log } u_{it}^2$, on peut réécrire (5.22) sous la forme :

$$\begin{cases} Z_t = -1.27e + h_t + \xi_t \\ h_t = \gamma + \text{diag}(\phi)h_{t-1} + v_t \end{cases} \quad (5.23)$$

où e est le vecteur dont toutes les composantes valent 1, γ est le vecteur de composantes γ_i , h_t le vecteur de composantes h_{it} et $\text{diag}(\phi)$ la matrice diagonale dont les éléments diagonaux sont les ϕ_i .

Le modèle (5.23) est, à nouveau, un modèle à espace d'états linéaire et les résultats du cas univarié se généralisent immédiatement.

On peut aussi songer à faire intervenir un vecteur h_t^* dont les composantes, ou facteurs, sont en nombre p inférieur à n ; on aboutit à un modèle du type :

$$\begin{cases} Z_t = -1.27e + \Theta h_t^* + \xi_t \\ h_t^* = \gamma^* + \text{diag}(\phi^*)h_{t-1}^* + v_t^* \end{cases} \quad (5.24)$$

Pour rendre le modèle identifiable on peut, par exemple dans le cas où $\text{diag}(\phi^*) = I$ (voir Harvey, Ruiz et Shephard [1992]), imposer que la $(p \times p)$ matrice carrée supérieure de la $(n \times p)$ matrice Θ soit triangulaire inférieure et que la matrice de variance-covariance de v_t^* soit l'identité. Les composantes de h_t^* peuvent être considérées comme des facteurs de volatilité.

CHAPITRE 6

INFÉRENCE FONDÉE SUR DES SIMULATIONS

par

Alain MONFORT

6.1 INTRODUCTION

Les simulations sont utilisées depuis longtemps pour évaluer les performances des méthodes d'estimation et de tests ou pour résoudre des problèmes numériques d'optimisation (algorithme E.M. simulé) et de calcul de lois a posteriori dans la cadre bayésien. Depuis peu⁽¹⁾ il a été proposé d'utiliser des simulations pour définir des fonctions critères permettant de dériver des procédures d'estimation et de test dans des contextes où les procédures classiques ne fonctionnent pas en raison de la complexité du modèle. Les modèles pour lesquels ces méthodes peuvent être utiles sont très variés (voir Gourieroux, Monfort et Renault [1992]) et, dans ce chapitre, on va se limiter à l'étude de modèles de « séries temporelles », c'est-à-dire sans variable exogène. On va considérer successivement la méthode des moments simulés, la méthode du maximum de vraisemblance simulée, la méthode du pseudo-maximum de vraisemblance simulée et les méthodes d'inférence indirecte.

⁽¹⁾ Lerman et Manski [1981], Van Praag et Hop [1987], Mc Fadden [1989], Pakes et Pollard [1989], Laroque et Salanié [1989], [1990], [1992], Duffie et Singleton [1989], Hajivassiliou et McFadden [1989], Borsch, Supan, Hajivassiliou et Keane [1990], Gourieroux et Monfort [1990], [1991a], [1991b], [1992], Gourieroux, Monfort et Renault [1992].

Le modèle de référence sera :

$$\begin{cases} y_t = r(y_{t-1}, u_t, \theta) \\ u_t = \phi(u_{t-1}, \varepsilon_t, \theta) \end{cases} \quad \theta \in \Theta \subset \mathbb{R}^p \quad (6.1)$$

y_t est un vecteur de variables endogènes observables et $\{\varepsilon_t\}$ un bruit blanc inobservable de loi connue G_0 ; cette dernière condition n'est pas restrictive dans un cadre paramétrique car on peut toujours supposer que les paramètres inconnus de la loi de ε_t ont été incorporés dans θ ; u_t s'interprète comme un terme d'erreur en général inobservable, suivant un processus markovien d'ordre 1. Le processus $\{y_t\}$ est supposé stationnaire.

On note $f(y_t/y_{t-1}; \theta)$ la densité conditionnelle de y_t sachant $y_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$ et on se place dans le cas où cette densité n'a pas de forme explicite et où on ne peut donc pas appliquer la méthode du maximum de vraisemblance.

6.2 LA MÉTHODE DES MOMENTS SIMULÉS

La méthode des moments simulés a été proposée par Mc Fadden [1989] pour des modèles de choix discrets sur données individuelles et adaptée au cas des modèles de séries temporelles par Duffie et Singleton [1989]. C'est ce dernier cas qui nous intéresse ici.

On considère une fonction multidimensionnelle $k(y_t, y_{t-1}, \dots, y_{t-r})$ et le moment empirique associé :

$$\bar{k}_T = \frac{1}{T} \sum_{t=r+1}^T k(y_t, \dots, y_{t-r}).$$

On suppose que θ est identifiable à partir de l'espérance de k . La méthode des moments simulés consiste à minimiser en θ la quantité :

$$\left[\bar{k}_T - \frac{1}{TH} \sum_{t=r+1}^{TH} k[\tilde{y}_t(\theta), \dots, \tilde{y}_{t-r}(\theta)] \right] \Omega \left[\bar{k}_T - \frac{1}{TH} \sum_{t=r+1}^{TH} k[\tilde{y}_t(\theta), \dots, \tilde{y}_{t-r}(\theta)] \right] \quad (6.2)$$

où Ω est une matrice symétrique définie positive et où les $\tilde{y}_t(\theta)$, $t = r + 1, \dots, TH$ sont des valeurs simulées obtenues par :

$$\begin{aligned} \tilde{y}_t(\theta) &= r[\tilde{y}_{t-1}(\theta), \tilde{u}_t(\theta), \theta] \quad t = 1, \dots, TH \\ \tilde{u}_t(\theta) &= \phi(\tilde{u}_{t-1}(\theta), \tilde{\varepsilon}_t, \theta) \end{aligned}$$

(avec $y_0(\theta) = y_0$ fixé), les $\tilde{\varepsilon}_t$ ayant été tirés successivement dans la loi (connue) de ε_t (u_0 est fixé). Il est important de noter que les tirages des $\tilde{\varepsilon}_t$ restent les mêmes quand θ varie dans la procédure de minimisation.

L'estimateur $\tilde{\theta}_T^H(\Omega)$ ainsi obtenu est convergent, asymptotiquement normal lorsque H est fixe et T tend vers l'infini. La matrice de variance-covariance asymptotique de $\sqrt{T}[\tilde{\theta}_T^H(\Omega) - \theta_0]$ est, en notant θ_0 la vraie valeur de θ :

$$\left(1 + \frac{1}{H}\right) \left[E_\theta \frac{\partial k'}{\partial \theta} \Omega E_{\theta_0} \frac{\partial k}{\partial \theta'} \right]^{-1} \times E_{\theta_0} \frac{\partial k'}{\partial \theta} \Omega V(\theta_0) \Omega E_{\theta_0} \frac{\partial k}{\partial \theta'} \left[E_\theta \frac{\partial k'}{\partial \theta} \Omega E_{\theta_0} \frac{\partial k}{\partial \theta'} \right]^{-1} \tag{6.1}$$

où $V(\theta_0)$ est la matrice de variance-covariance asymptotique de

$$\frac{1}{\sqrt{T}} \sum_{t=r+1}^T [k(y_t, \dots, y_{t-r}) - E_{\theta_0} k],$$

à savoir :

$$\sum_{h=-\infty}^{\infty} \Gamma_h,$$

avec $\Gamma_h = \text{cov}(k_{t-h}, k_t)$ et $k_t = k(y_t, \dots, y_{t-r})$ et où les diverses matrices sont évaluées en $\theta = \theta_0$.

L'estimateur $\tilde{\theta}_T(\Omega)$ dépend de Ω et on peut montrer que l'estimateur optimal est obtenu en prenant $\Omega = \hat{V}^{-1}$, où \hat{V} est un estimateur convergent de $V(\theta_0)$. Un tel estimateur peut être obtenu par la méthode de Newey-West [1987] :

$$\hat{V} = \hat{\Gamma}_0 + \sum_{j=1}^{R_T} \left(1 - \frac{j}{R_T + 1}\right) (\hat{\Gamma}_j + \hat{\Gamma}'_j) \tag{6.4}$$

avec :

$$\hat{\Gamma}_j = \frac{1}{T} \sum_{t=j+r+1}^T k_{t-j} k'_t$$

où R_T est une suite d'entiers tendant vers l'infini plus lentement que $T^{1/4}$.

L'estimateur optimal ainsi obtenu est noté $\hat{\theta}_T^H$ et la matrice de variance-covariance asymptotique de $\sqrt{T}(\hat{\theta}_T^H - \theta_0)$ est :

$$\left(1 + \frac{1}{H}\right) \left[E_{\theta_0} \frac{\partial k'}{\partial \theta} V(\theta_0)^{-1} E_{\theta_0} \frac{\partial k}{\partial \theta'} \right]^{-1}. \tag{6.5}$$

Cette matrice est donc égale, au facteur $1 + \frac{1}{H}$ près, à la matrice de variance-covariance asymptotique de l'estimateur des moments, qui serait obtenu en minimisant en θ :

$$[\bar{k}_T - E_\theta k]' \hat{V}^{-1} [\bar{k}_T - E_\theta k]. \tag{6.6}$$

Naturellement cet estimateur n'est pas utilisable si la fonction de θ , $E_\theta k$, n'est pas connue, alors que l'estimateur des moments simulés est toujours utilisable et c'est ce qui fait son intérêt.

L'efficacité asymptotique des moments est $\frac{H}{1+H}$, soit 50% pour $H = 1$, 90% pour $H = 9$.

La méthode précédente a été proposée par Duffie et Singleton [1989] pour l'estimation des paramètres du processus de diffusion défini par :

$$dy_t = \mu(y_t, \theta)dt + \sigma(y_t, \theta)dB_t \quad t \in \mathbb{R}^+$$

où $\{B_t, t \in \mathbb{R}^+\}$ est un mouvement brownien. L'idée consiste à approximer le processus $\{y_t, t \in \mathbb{R}^+\}$ par un processus discret de pas $1/n$ obtenu, par exemple, par la méthode d'Euler :

$$\tilde{y}_\tau^{(n)} = \tilde{y}_{\tau-1}^{(n)} + \frac{1}{n} \mu(\tilde{y}_{\tau-1}^{(n)}, \theta) + \frac{1}{\sqrt{n}} \sigma(\tilde{y}_{\tau-1}^{(n)}, \theta) \tilde{\varepsilon}_\tau \quad (6.7)$$

où $\tilde{\varepsilon}_\tau$ est un bruit blanc gaussien de variance 1. Ce processus est observé toutes les n périodes et, par conséquent, les méthodes classiques, comme le maximum de vraisemblance, sont difficilement utilisables. En revanche, comme le processus $\tilde{y}_\tau^{(n)}$ peut être facilement simulé, on peut obtenir simplement les simulations $\tilde{y}_t(\theta)t = 1, \dots, TH$ nécessaires pour la mise en oeuvre de la méthode des moments simulés.

6.3 LA MÉTHODE DU MAXIMUM DE VRAISEMBLANCE SIMULÉE

Plusieurs versions de la méthode du maximum de vraisemblance simulée ont été proposées (voir Gourieroux et Monfort (1991b) pour une synthèse). On présente ici brièvement la méthode fondée sur un conditionnement. Dans ce type de méthode on suppose que l'on peut trouver un bruit blanc conditionnant $\{v_t\}$, où v_t est indépendant de \underline{y}_{t-1} de loi connue P^V et tel que la densité conditionnelle $f^*(y_t/\underline{y}_{t-1}, v_t, \theta)$ ait une forme explicite.

Dans ce cas, on a :

$f(y_t/\underline{y}_{t-1}, \theta) = \int f^*(y_t/\underline{y}_{t-1}, v, \theta)dP^V(v)$ et donc $f(y_t/\underline{y}_{t-1}, \theta)$ peut être approximée par :

$$\frac{1}{H} \sum_{h=1}^H f^*(y_t/\underline{y}_{t-1}, v_{th}, \theta) \quad (6.8)$$

où les v_{th} sont tirés indépendamment dans P^V .

La méthode du maximum de vraisemblance simulée consiste alors à maximiser en θ :

$$\sum_{t=1}^T \text{Log} \left[\frac{1}{H} \sum_{h=1}^H f^*(y_t/\underline{y}_{t-1}, v_{th}, \theta) \right]. \quad (6.9)$$

On peut montrer que lorsque H et T tendent vers l'infini, de façon que $\frac{\sqrt{T}}{H}$ tende vers zéro, l'estimateur ainsi obtenu a les mêmes propriétés asymptotiques que l'estimateur du maximum de vraisemblance.

Si on reprend l'exemple précédent de l'approximation d'Euler d'un processus de diffusion et si on considère, pour simplifier, le cas $n = 2$ on a, avec des notations évidentes :

$$\begin{aligned} y_t &= F\left(\tilde{y}_{2t-1}^{(2)}, \tilde{\varepsilon}_{2t}, \theta\right) \\ &= F\left[F(y_{t-1}, \tilde{\varepsilon}_{2t-1}, \theta), \tilde{\varepsilon}_{2t}, \theta\right]. \end{aligned}$$

Conditionnellement à y_{t-1} , et $\tilde{\varepsilon}_{2t-1}$ (qui joue le rôle de v_t), la loi de y_t est une loi normale de moyenne

$$F(y_{t-1}, \tilde{\varepsilon}_{2t-1}, \theta) + \frac{1}{n} \mu [F(y_{t-1}, \tilde{\varepsilon}_{2t-1}, \theta), \theta]$$

et de variance

$$\frac{1}{n} \sigma^2 [F(y_{t-1}, \tilde{\varepsilon}_{2t-1}, \theta), \theta].$$

On est donc bien dans les conditions d'application de la méthode du maximum de vraisemblance simulée, si on suppose que l'approximation d'Euler est « bien spécifiée », c'est-à-dire si on néglige cette approximation.

6.4 LA MÉTHODE DU PSEUDO-MAXIMUM DE VRAISEMBLANCE SIMULÉE

Les méthodes du pseudo-maximum de vraisemblance ont été proposées par Gourieroux, Monfort et Trognon [1984] dans un cadre statique, mais elles peuvent facilement s'adapter au cadre dynamique qui nous intéresse ici. Notons $m(\theta)$ l'espérance mathématique de y_t et $v(\theta)$ sa matrice de variance-covariance. Pour estimer θ , on peut, en théorie, utiliser une méthode du pseudo-maximum de vraisemblance fondée sur ces moments marginaux et sur une famille exponentielle linéaire ou quadratique. On se bornera ici à décrire les méthodes du premier (PMV1) et du deuxième ordre (PMV2) fondées sur la famille normale.

La méthode du PMV1 se réduit, dans ce cas particulier, à la méthode des moindres carrés non linéaires et est fondée sur la minimisation de :

$$L_{1T}(\theta) = \sum_{t=1}^T \ell_{1t}(\theta) = \frac{1}{2} \sum_{t=1}^T [y_t - m(\theta)]' \Omega [y_t - m(\theta)] \quad (6.10)$$

où Ω est une matrice symétrique définie positive donnée (par exemple $\Omega = \text{Id}$).

La méthode du PMV2 revient, dans le cas où la famille de loi utilisée est la famille normale, à minimiser :

$$\begin{aligned} L_{2T}(\theta) &= \sum_{t=1}^T l_{2t}(\theta) \\ &= \frac{1}{2} \sum_{t=1}^T [\text{Log det } v(\theta) + (y_t - m(\theta))' v^{-1}(\theta) (y_t - m(\theta))] \end{aligned} \quad (6.11)$$

On peut montrer que, sous certaines hypothèses de régularité, ces estimateurs sont convergents et asymptotiquement normaux et de matrice de variance-covariance asymptotique :

$$J_{i_0}^{-1} I_{i_0} J_{i_0}^{-1} \quad i = 1, 2 \quad (6.12)$$

avec :

$$\begin{aligned} I_{i_0} &= \lim_{T \rightarrow \infty} \frac{1}{T} \text{V}_0 \frac{\partial L_{iT}}{\partial \theta}(\theta_0) \\ J_{i_0} &= - \lim_{T \rightarrow \infty} \text{E}_0 \frac{1}{T} \frac{\partial^2 L_{iT}}{\partial \theta \partial \theta'}(\theta_0) \quad i = 1, 2. \end{aligned}$$

La matrice I_{i_0} peut à nouveau être estimée par la méthode de Newey-West [1987] :

$$\begin{aligned} \hat{I}_{iT} &= \frac{1}{T} \sum_{i=1}^T \frac{\partial \ell_{it}}{\partial \theta}(\hat{\theta}_{it}) \frac{\partial \ell_{it}}{\partial \theta'}(\hat{\theta}_{it}) + \sum_{j=1}^{R_T} \left(1 - \frac{j}{R_T + 1}\right) \\ &\quad \times \sum_{t=j+1}^T \left(\frac{\partial \ell_{it}}{\partial \theta}(\hat{\theta}_{iT}) \frac{\partial \ell_{i,t-j}}{\partial \theta'}(\hat{\theta}_{iT}) + \frac{\partial \ell_{i,t-j}}{\partial \theta}(\hat{\theta}_{iT}) \frac{\partial \ell_{i,t}}{\partial \theta'}(\hat{\theta}_{iT}) \right) \end{aligned}$$

où $\hat{\theta}_{iT}$ et $i = 1, 2$ sont les estimateurs du PMV1 et du PMV2 respectivement. La matrice J_{i_0} peut être estimée par :

$$\hat{J}_{iT} = - \frac{1}{T} \frac{\partial^2 L_{iT}}{\partial \theta \partial \theta'}(\hat{\theta}_{iT}).$$

On peut aussi trouver un estimateur de J_{i_0} ne faisant intervenir que des dérivées premières (voir annexe 5 de Gourieroux, Monfort et Trognon [1984]).

Parmi les hypothèses requises pour que les résultats précédents soient valables, il y a évidemment des hypothèses d'identification : pour la méthode PMV1 θ doit être identifiable au premier ordre, c'est-à-dire à partir de $m(\theta)$ et pour la méthode de PMV2 θ doit être identifiable au deuxième ordre, c'est-à-dire à partir de $m(\theta)$ et $v(\theta)$. Lorsque ces hypothèses d'identification ne sont pas satisfaites on peut soit se contenter de l'estimation de fonctions identifiables, soit changer de vecteur y_t de façon à retrouver l'identification ; en particulier on peut

songer à transformer y_t en un vecteur dont les composantes sont fonctions de $y_t, y_{t-1}, y_t^2, y_{t-1}^2, y_t y_{t-1}$ etc.

Le problème de base rencontré dans la mise en œuvre de telles méthodes est l'absence de forme explicite pour $m(\theta)$ et $v(\theta)$; il est donc naturel de songer à des méthodes du pseudo-maximum de vraisemblance simulées dans lesquelles $m(\theta)$ et $v(\theta)$ sont remplacés par des approximations fondées sur des simulations.

En fait, ces méthodes du PMV simulées peuvent s'appliquer à des modèles plus complexes que (6.1) dans lesquels apparaissent des variables endogènes inobservables. Ces modèles peuvent s'écrire :

$$\begin{cases} y_t = r(y_{t-1}, y_{t-1}^*, u_t, \theta) \\ y_t^* = r^*(y_{t-1}, y_{t-1}^*, u_t, \theta) \\ u_t = \phi(u_{t-1}, \varepsilon_t, \theta). \end{cases} \quad (6.13)$$

On voit que pour de tels modèles la vraisemblance est très complexe puisqu'elle s'obtient à partir de la densité de $y_1, \dots, y_T, y_1^*, \dots, y_T^*$ en intégrant par rapport à y_1^*, \dots, y_T^* . En revanche il est facile de simuler ce modèle pour des valeurs données de θ, y_0, y_0^*, u_0 ; il suffit en effet de tirer indépendamment TH valeurs de ε_t dans sa loi marginale : $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_{TH}$ et de calculer les $\tilde{y}_t^{(\theta)}, t = 1, \dots, TH$ par :

$$\begin{aligned} \tilde{y}_t(\theta) &= r(\tilde{y}_{t-1}(\theta), \tilde{y}_{t-1}^*(\theta), \tilde{u}_t(\theta), \theta) \\ \tilde{y}_t^*(\theta) &= r^*(\tilde{y}_{t-1}(\theta), \tilde{y}_{t-1}^*(\theta), \tilde{u}_t(\theta), \theta) \\ \tilde{u}_t(\theta) &= \phi(\tilde{u}_{t-1}(\theta), \tilde{\varepsilon}_t, \theta) \end{aligned}$$

On peut alors calculer des approximations de $m(\theta)$ et $v(\theta)$:

$$\begin{aligned} m_{TH}(\theta) &= \frac{1}{TH} \sum_{t=1}^{TH} \tilde{y}_t(\theta) \\ v_{TH}(\theta) &= \frac{1}{TH} \sum_{t=1}^{TH} [\tilde{y}_t(\theta) - m_{TH}(\theta)] [\tilde{y}_t(\theta) - m_{TH}(\theta)]'. \end{aligned}$$

Si on remplace $m(\theta)$ et $v(\theta)$ par ces approximations dans (6.10) et (6.11) on obtient les méthodes du PMV1 et PMV2 simulées. Les propriétés des méthodes du pseudo-maximum de vraisemblance simulées ont été montrées dans le cas statique par Laroque et Salanié ([1989], [1990]) et elles se généralisent au cas dynamique stationnaire ; le résultat principal est la convergence des estimateurs lorsque T et H tendent vers l'infini et la convergence en loi vers la même loi asymptotique que les estimateurs du PMV correspondant pourvu que T et H tendent vers l'infini et $\frac{\sqrt{T}}{H}$ tende vers zéro.

Ce type de méthode a été appliquée dans le cas statique et dans le cas dynamique par Laroque et Salanié ([1989], [1992]) à des modèles de déséquilibre. Dans le domaine financier on pourrait penser à les utiliser pour estimer des modèles à volatilité stochastique continue (voir 5.4) puisque, dans ces modèles, il existe une variable endogène inobservable, à savoir σ_t .

6.5 MÉTHODES D'INFÉRENCE INDIRECTE

6.5.1 Présentation des méthodes

On considère à nouveau le modèle (6.1); on suppose que ce modèle est complexe et ne permet pas une estimation directe simple.

Les méthodes d'inférence indirecte (voir Gourieroux, Monfort et Renault [1992]) font intervenir une fonction critère $Q_T(\underline{y}_T, \beta)$, où $\underline{y}_T = (y_1, \dots, y_T)$ et β est un paramètre auxiliaire.

On note $\hat{\beta}_T$ la solution du problème de maximisation de Q_T en β :

$$\hat{\beta}_T = \underset{\beta}{\text{Argmax}} Q_T(\underline{y}_T, \beta). \quad (6.14)$$

On suppose que, lorsque T tend vers l'infini, Q_T converge (presque sûrement uniformément en β) vers une limite déterministe $Q_\infty(\theta_0, \beta)$, où θ_0 est la vraie valeur de θ . On note :

$$\begin{cases} \beta_0 = \underset{\beta}{\text{Argmax}} Q_\infty(\theta_0, \beta) \\ b(\theta) = \underset{\beta}{\text{Argmax}} Q_\infty(\theta, \beta). \end{cases} \quad (6.15)$$

b est appelée fonction de lien et on suppose que la fonction b est injective. Si cette fonction était connue on pourrait estimer θ en résolvant $b(\theta) = \hat{\beta}_T$ mais comme cette fonction est, en général, inconnue, on va avoir recours à des simulations. Pour chaque valeur de θ on peut simuler, comme on l'a vu, TH valeurs de y , notées $\tilde{y}_{TH}(\theta)$. A partir de ces simulations on peut calculer :

$$\tilde{\beta}_{TH}(\theta) = \underset{\beta}{\text{Argmax}} Q_T(\tilde{y}_{TH}(\theta), \beta). \quad (6.16)$$

Un estimateur indirect de θ est alors obtenu par :

$$\tilde{\theta}_{TH}(\Omega) = \underset{\theta}{\text{Argmin}} \left(\hat{\beta}_T - \tilde{\beta}_{TH}(\theta) \right)' \hat{\Omega}_T \left(\hat{\beta}_T - \tilde{\beta}_{TH}(\theta) \right) \quad (6.17)$$

où $\hat{\Omega}_T$ est une suite de matrices convergeant vers Ω symétrique définie positive.

Sous des hypothèses de régularité classiques, en particulier que $\partial b / \partial \theta$ a pour rang la dimension de θ , on peut montrer que, lorsque H est fixe et T tend

vers l'infini, $\tilde{\theta}_{TH}(\Omega)$ est convergent et que

$$\sqrt{T} \left[\tilde{\theta}_{TH}(\Omega) - \theta_0 \right]$$

converge en loi vers $N[0, W_H(\Omega)]$, avec :

$$W_H(\Omega) = \left(1 + \frac{1}{H} \right) \left(\frac{\partial b'}{\partial \theta} \Omega \frac{\partial b}{\partial \theta'} \right)^{-1} \frac{\partial b'}{\partial \theta} \Omega J_0^{-1} I_0 J_0^{-1} \frac{\partial b}{\partial \theta'} \left(\frac{\partial b'}{\partial \theta} \Omega \frac{\partial b}{\partial \theta'} \right)^{-1} \quad (6.18)$$

avec :

$$I_0 = \lim_{T \rightarrow \infty} T V_0 \left(\frac{\partial Q_T}{\partial \beta} (\underline{y}_T, \beta_0) \right)$$

$$J_0 = - \frac{\partial^2 Q_\infty}{\partial \beta \partial \beta'} (\theta_0, \beta_0)$$

et où toutes les matrices sont évaluées en $\theta = \theta_0$.

Cet estimateur $\tilde{\theta}_{TH}(\Omega)$ dépend de Ω et il est facile de voir que l'estimateur le plus efficace asymptotiquement dans cette classe est obtenu pour :

$$\Omega = \Omega^* = J_0 I_0^{-1} J_0 \quad (6.19)$$

et on note $\hat{\theta}_{TH} = \tilde{\theta}_{TH}(\Omega^*)$.

La matrice de variance-covariance asymptotique de $\sqrt{T}(\hat{\theta}_{TH} - \theta_0)$ est alors :

$$W_H^* = \left(1 + \frac{1}{H} \right) \left[\frac{\partial b'}{\partial \theta} J_0 I_0^{-1} J_0 \frac{\partial b}{\partial \theta'} \right]^{-1} \quad (6.20)$$

$$= \left(1 + \frac{1}{H} \right) \left[\frac{\partial^2 Q_\infty}{\partial \theta \partial \beta'} I_0^{-1} \frac{\partial^2 Q_\infty}{\partial \beta \partial \theta'} \right]^{-1}.$$

6.5.2 Remarques

i) Remarque 1

Comme on l'a déjà mentionné, il est important de souligner que dans le calcul de $\tilde{\theta}_{TH}(\Omega)$ [voir (6.17)], les valeurs des simulations $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_{TH}$ doivent rester les mêmes lorsque θ varie.

ii) Remarque 2

Le calcul de $\tilde{\theta}_{TH}(\Omega)$ ne nécessite le calcul de $\tilde{\beta}_{TH}(\theta)$ que pour les valeurs de θ utilisées par l'algorithme de minimisation employé.

iii) Remarque 3

Le calcul de $\hat{\theta}_{TH}$ nécessite le calcul d'un estimateur convergent $\hat{\Omega}_T^*$ de Ω^* qui, à son tour, nécessite un estimateur convergent de θ . Il faut donc procéder en deux étapes : dans une première étape on calcule un estimateur convergent de θ , à savoir $\tilde{\theta}_{TH}(\Omega)$ avec Ω fixée (par exemple $\Omega = \text{Id}$), et dans une deuxième étape on calcule Ω_T^* et $\hat{\theta}_{TH}$.

iv) Remarque 4

Si les dimensions de θ et β sont égales on obtient directement l'estimateur optimal $\tilde{\theta}_{TH}$ en prenant pour Ω n'importe quelle matrice symétrique définie positive.

v) Remarque 5

On vérifie facilement que si on prend comme fonction critère

$$Q_T = -\frac{1}{2}(\bar{k}_T - \beta)'(\bar{k}_T - \beta)$$

où $\bar{k}_T = \frac{1}{T} \sum_{t=r+1}^T k(y_t, \dots, y_{t-r})$ est défini comme dans le paragraphe 6.2, on retrouve la méthode des moments simulés. Celle-ci est donc une méthode d'inférence indirecte particulière.

vi) Remarque 6

On peut montrer que la famille des estimateurs $\tilde{\theta}_{TH}(\Omega)$ a les mêmes propriétés asymptotiques que la famille $\theta_{TH}^*(W)$ où $\theta_{TH}^*(W)$ est obtenu en minimisant θ :

$$\frac{\partial Q_T}{\partial \beta'} \left(\tilde{y}_{TH}(\theta), \hat{\beta}_T \right) W \frac{\partial Q_T}{\partial \beta} \left(\tilde{y}_{TH}(\theta), \hat{\beta}_T \right).$$

Si les dérivés $(\partial Q_T / \partial \beta)$ sont facilement explicitables, cette méthode a l'avantage de ne pas nécessiter des optimisations emboîtées. La matrice W optimale est $W^* = I_0^{-1}$.

Les procédures d'estimation que l'on vient de décrire peuvent être complétées par des procédures de test.

Tout d'abord un test de spécification est obtenu à partir de la valeur optimale de la fonction ayant servi à calculer $\hat{\theta}_{TH}$; plus précisément la statistique :

$$\xi_T = \frac{TH}{1+H} \min_{\theta} \left[\hat{\beta}_T - \tilde{\beta}_{TH}(\theta) \right]' \Omega_T^* \left[\hat{\beta}_T - \tilde{\beta}_{TH}(\theta) \right] \quad (6.21)$$

suit, asymptotiquement, lorsque T tend vers l'infini et H est fixe, sous l'hypothèse que le modèle (6.1) est bien spécifié, la loi $\chi^2(q-p)$, où q est la dimension de β et p celle de θ . Donc, si q est strictement supérieur à p , on en déduit un

test de spécification de niveau asymptotique α en prenant comme région critique $\{\xi_T \geq \chi^2_{1-\alpha}(q-p)\}$.

Si on s'intéresse maintenant à des tests de significativité, c'est-à-dire à des contraintes sur θ , on peut montrer (voir Gourieroux, Monfort et Renault [1992]) que les tests du score et de Wald conservent leurs propriétés asymptotiques usuelles malgré la présence de simulations ; il en est de même du test fondé sur la différence des valeurs optimales de la fonction objectif (généralisant le test du rapport de vraisemblance) à condition de prendre comme fonction objectif la fonction « optimale », c'est-à-dire celle apparaissant dans (6.17) en remplaçant $\hat{\Omega}_T$ par $\hat{\Omega}_T^*$.

Les domaines d'application des méthodes indirectes sont très variés (voir Gourieroux, Monfort et Renault [1992]). On se bornera ici à décrire quelques possibilités d'applications à des modèles financiers.

Une première grande catégorie d'applications est celle dans laquelle la fonction critère est obtenue en approximant la fonction de vraisemblance exacte. Un exemple est fourni par l'utilisation du filtre de Kalman étendu (voir 5.2.5) pour approximer la fonction de vraisemblance d'un modèle GARCH à facteurs. Dans ce type d'applications, la méthode d'estimation indirecte peut être vue comme un moyen de corriger automatiquement le biais d'estimation qui apparaîtrait si on se bornait à maximiser la vraisemblance approchée.

Une deuxième grande catégorie d'applications est celle dans laquelle la fonction critère est la fonction de vraisemblance exacte d'un modèle approché. Un premier exemple de ce type est obtenu en considérant l'approximation d'un facteur GARCH par une chaîne de Markov (voir 5.2.5) et en utilisant éventuellement l'algorithme E.M. pour maximiser la vraisemblance du modèle approché. Nous allons fournir deux autres exemples appartenant à cette deuxième catégorie.

6.5.3 Estimation des paramètres d'une équation différentielle stochastique (EDS)

On considère un processus de diffusion défini par l'EDS :

$$dy_t = g(\theta, y_t)dt + h(\theta, y_t)dB_t \tag{6.22}$$

où B_t est un mouvement brownien (de variance égale à 1).

On suppose que le processus est observé aux dates $1, \dots, T$. La distribution exacte de y_1, \dots, y_T est, en général, inconnue (sauf pour des fonctions g et h particulières) et on ne peut donc pas appliquer la méthode du maximum de vraisemblance. On peut cependant fonder une méthode d'estimation indirecte sur la vraisemblance du modèle discrétisé approché obtenu à partir de l'EDS suivi

par y_t ou, plus généralement, par $k(y_t)$. En utilisant la formule d'Ito on a :

$$dk(y_t) = \left[\frac{\partial k}{\partial y}(y_t)g(\theta, y_t) + \frac{1}{2} \frac{\partial^2 k}{\partial y^2}(y_t)h^2(\theta, y_t) \right] dt + \frac{\partial k}{\partial y}(y_t)h(\theta, y_t)dB_t \quad (6.23)$$

Une discrétisation est :

$$k(y_t) = k(y_{t-1}) + \frac{\partial k}{\partial y}(y_{t-1})g(\beta, y_{t-1}) + \frac{1}{2} \frac{\partial^2 k}{\partial y^2}(y_{t-1})h^2(\beta, y_{t-1}) + \frac{\partial k}{\partial y}(y_{t-1})h(\beta, y_{t-1})\varepsilon_t \quad (6.24)$$

où ε_t est un bruit blanc de variance égale à 1. On obtient donc un modèle autorégressif d'ordre 1 non linéaire avec effet ARCH dont la vraisemblance peut être calculée facilement et on peut appliquer une méthode d'estimation indirecte.

6.5.4 Estimation des paramètres d'un modèle à volatilité stochastique en temps continu

Considérons, par exemple, le modèle :

$$\begin{cases} dy_t = \mu y_t dt + \sigma_t y_t dB_{1t} \\ d \text{Log } \sigma_t = a(b - \text{Log } \sigma_t) dt + cd B_{2t} \end{cases} \quad (6.25)$$

où (B_{1t}, B_{2t}) est un brownien bivarié de matrice de variance-covariance :

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Si la volatilité σ_t était fixe, y_t suivrait un mouvement brownien géométrique, mais ici on suppose que σ_t est inobservable et que $\text{Log } \sigma_t$ suit un processus d'Ornstein-Uhlenbeck. Comme dans l'exemple précédent on peut considérer le modèle approché obtenu en discrétisant avec un pas égal à celui des observations. Cependant, comme σ_t est inobservable, on obtient un modèle espace d'états non linéaire du type :

$$\begin{cases} \text{Log } y_t = \lambda + \text{Log } y_{t-1} + \sigma_t \varepsilon_{1t} \\ \text{Log } \sigma_t = d_0 + d_1 \text{Log } \sigma_{t-1} + d_2 \varepsilon_{2t} \end{cases} \quad (6.26)$$

où $(\varepsilon_{1t}, \varepsilon_{2t})$ est un bruit blanc bivarié de matrice de variance-covariance :

$$\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}.$$

Un estimateur convergent de λ est

$$\hat{\lambda}_T = \frac{1}{T} \sum_{t=1}^T \text{Log} \frac{y_t}{y_{t-1}}$$

et on peut donc prendre comme fonction critère la pseudo log-vraisemblance du modèle approché (voir Harvey, Ruiz et Shephard [1992]) :

$$\begin{cases} \text{Log} \left[\text{Log} \frac{y_t}{y_{t-1}} - \hat{\lambda}_t \right]^2 = \frac{1}{2} h_t + \text{Log} \varepsilon_{1t}^2 \\ h_t = d_0 + d_1 h_{t-1} + d_2 \varepsilon_{2t} \end{cases} \quad (6.27)$$

avec $h = \text{Log} \sigma_t$.

$\text{Log} \varepsilon_{1t}^2$ est de moyenne $-1,27$ et de variance 4.93 et si on pose $\xi = \text{Log} \varepsilon_{1t}^2 + 1.27$ on obtient un bruit blanc centré de variance 4.93 . Le modèle approché s'écrit alors :

$$\begin{cases} \text{Log} \left[\text{Log} \frac{y_t}{y_{t-1}} - \hat{\lambda}_T \right]^2 = -1,27 + \frac{1}{2} h_t + \xi_t \\ h_t = d_0 + d_1 h_{t-1} + d_2 \varepsilon_{2t} \end{cases} \quad (6.28)$$

Ce modèle est sous forme espace d'états linéaire et on peut donc utiliser le filtre de Kalman standard pour calculer la pseudo-vraisemblance associée au cas normal (et fondée sur la régression linéaire par rapport au passé au lieu de l'espérance conditionnelle). Cette fonction pourra être utilisée comme fonction critère et, ici encore, la deuxième étape de la méthode d'inférence indirecte peut être vue comme une méthode de redressement automatique des biais asymptotiques dus aux approximations.

6.5.5 Applications numériques

On va considérer deux applications au cas des équations différentielles stochastiques du type (6.22) et, de façon à évaluer l'efficacité de la méthode, on va considérer deux cas où on connaît la discrétisation exacte du modèle en temps continu et où on peut donc calculer les estimateurs du maximum de vraisemblance.

i) Cas du mouvement brownien géométrique

L'EDS est :

$$dy_t = \mu y_t dt + \sigma y_t dB_t$$

La discrétisation exacte est :

$$\text{Log } y_t = \text{Log } y_{t-1} + \mu - \frac{\sigma^2}{2} + \sigma \varepsilon_t$$

où ε_t est un bruit blanc gaussien centré réduit.

La discrétisation approchée est :

$$y_t = (1 + \mu)y_{t-1} + \sigma y_{t-1} \varepsilon_t.$$

L'estimateur indirect (EI) considéré est fondé sur la vraisemblance du modèle discrétisé approché et sur des simulations dans ce même modèle en prenant comme unité de temps le dixième de l'unité de temps pour les observations. On tire 150 observations dans le modèle discret exact avec $\mu = 0.2$, $\sigma = 0.5$, $y_0 = 10$ et on estime les paramètres μ et σ à partir d'une simulation ($H = 1$). Ces opérations sont répétées deux cents fois afin d'obtenir des distributions empiriques des estimateurs.

On estime également les paramètres par la méthode du maximum de vraisemblance exacte (MVE) et par la méthode du maximum de vraisemblance approchée (MVA). Les résultats sont résumés dans le tableau 6.1 et les figures 6.1 et 6.2. On voit que l'estimateur indirect est nettement moins biaisé que l'estimateur du maximum de vraisemblance approché et que son erreur quadratique moyenne est aussi nettement plus faible (pour σ) ou comparable (pour μ).

Tableau 6.1: Mouvement brownien

estimateur	moyenne	biais moyen	écart-type	racine de l'EQM
MVE				
μ	0,203	0,003	0,047	0,047
σ	0,496	-0,004	0,028	0,028
EI				
μ	0,208	0,008	0,068	0,069
σ	0,517	0,017	0,058	0,061
MVA				
μ	0,226	0,026	0,059	0,064
σ	0,646	0,146	0,072	0,163

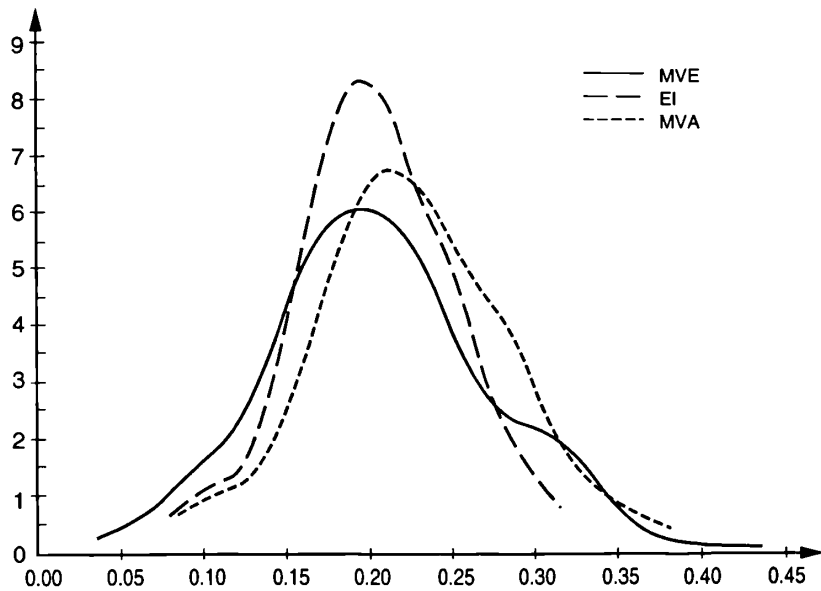


Figure 6.1: Brownien géométrique : densité des estimateurs de μ

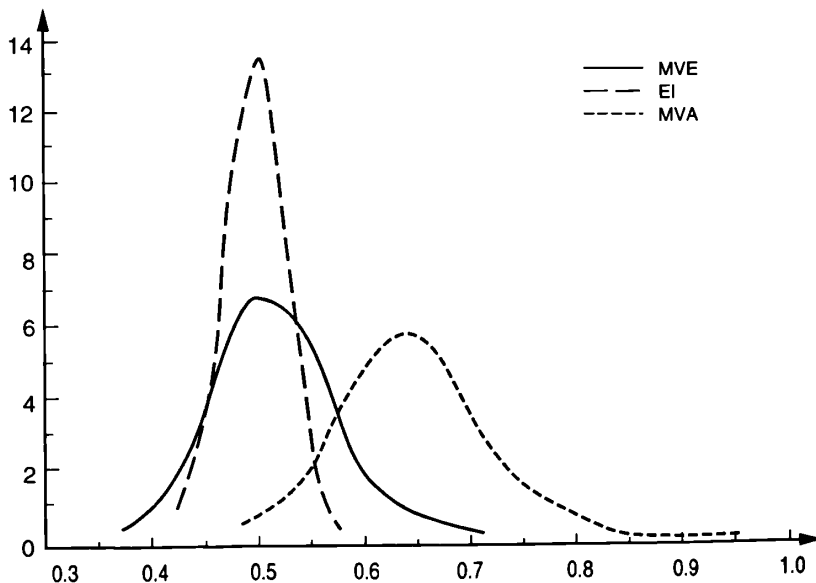


Figure 6.2: Brownien géométrique : densité des estimateurs de σ

ii) Cas du processus d'Ornstein-Uhlenbeck

L'EDS est :

$$dy_t = k(a - y_t)dt + \sigma dB_t.$$

La discrétisation exacte est :

$$y_t = a(1 - e^{-k}) + e^{-k}y_{t-1} + \sigma \left[\frac{(1 - e^{-2k})}{2k} \right]^{\frac{1}{2}} \varepsilon_t.$$

La discrétisation approchée est :

$$y_t = (1 - k)y_{t-1} + ka + \sigma \varepsilon_t.$$

Les valeurs numériques sont $k = 0.8$, $a = 0.1$, $\sigma = 0.06$, $y_0 = 0$, $H = 1$; on tire 250 observations et on répète 200 fois les opérations. Les résultats sont résumés dans le tableau 6.2 et les figures 6.3, 6.4 et 6.5. Il se trouve que, par coïncidence, les estimateurs indirects et approchés de a sont identiques ; pour l'estimation de k et σ l'estimateur indirect est nettement supérieur à l'estimation du maximum de vraisemblance approché. On note même que le support de l'estimateur du maximum de vraisemblance approché de σ est pratiquement disjoint du support des autres estimateurs et que cette méthode fait apparaître, à tort, une bonne précision.

Tableau 6.2: Processus d'Ornstein Uhlenbeck

estimateur	moyenne	biais moyen	écart-type	racine de l'EQM
MVE				
k	0,834	0,034	0,141	0,145
a	0,101	0,001	0,004	0,004
σ	0,048	-0,012	0,003	0,012
EI				
k	0,781	-0,019	0,180	0,181
a	0,100	0,000	0,005	0,005
σ	0,045	-0,015	0,010	0,018
MVA				
k	0,561	-0,239	0,060	0,246
q	0,101	0,001	0,004	0,004
σ	0,034	-0,026	0,002	0,026

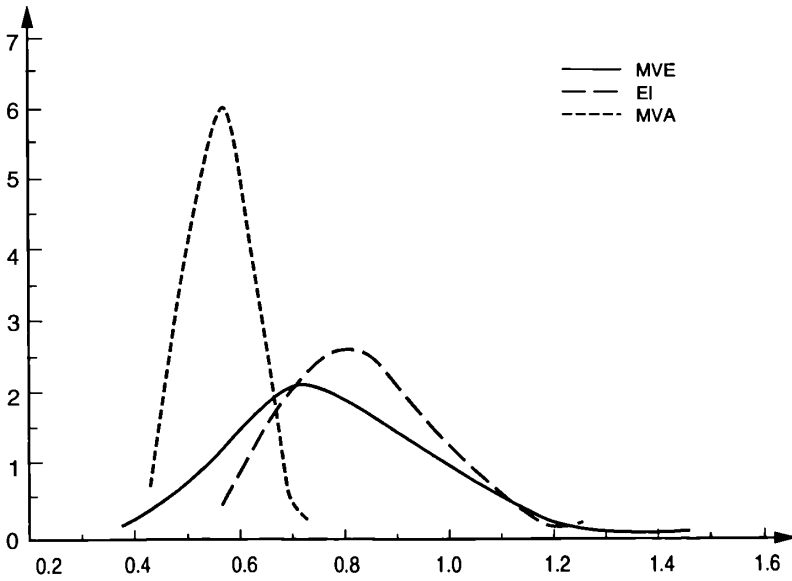


Figure 6.3: Processus d'Ornstein Uhlenbeck : densité des estimateurs de k

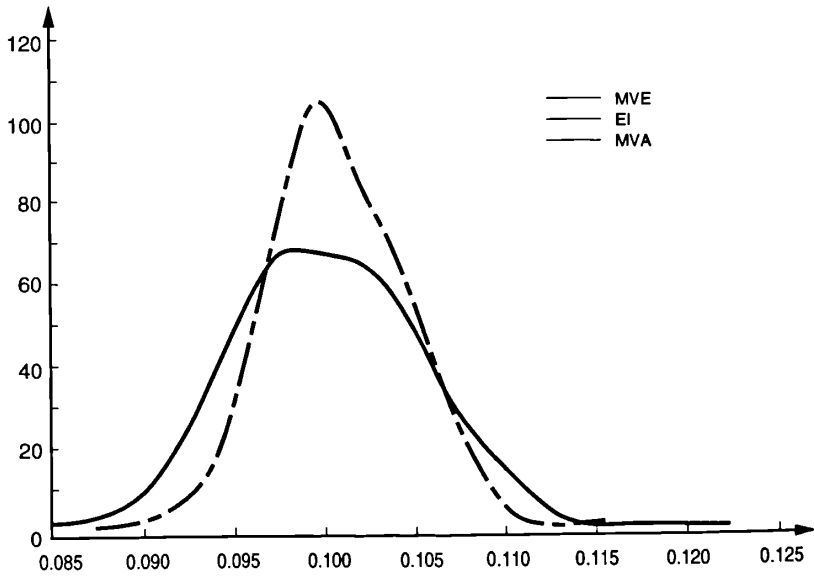


Figure 6.4: Processus d'Ornstein Uhlenbeck : densité des estimateurs de α

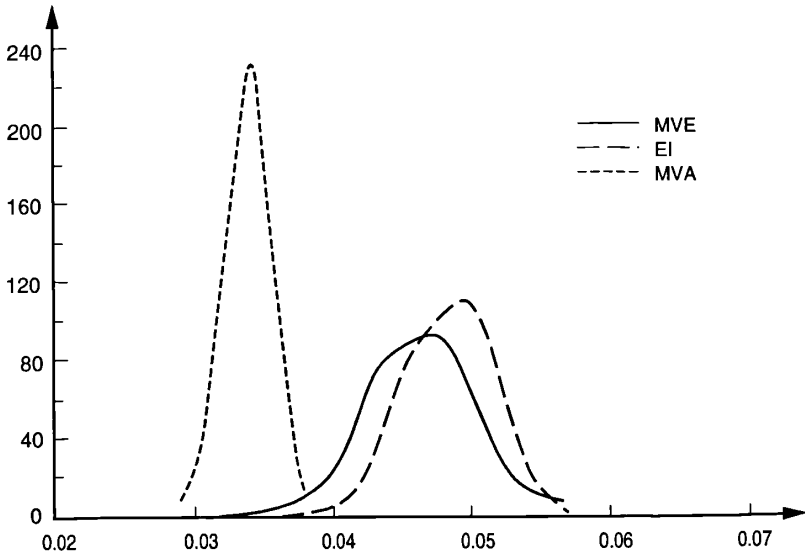


Figure 6.5: Processus d'Ornstein Uhlenbeck: densité des estimateurs de σ

CHAPITRE 7

MODÈLES A FACTEURS EN FINANCE

par

Sergio PASTORELLO et Eric RENAULT

7.1 INTRODUCTION

Il s'agit dans ce chapitre d'étudier les interactions entre la théorie financière (modèles d'évaluation d'actifs financiers) et la théorie statistique (modèles dynamiques à facteurs).

La notion de « modèle à facteur » (ou modèle à index ou à coefficients bêtas, . . .) est ancienne en Finance. Depuis les travaux fondateurs de Markowitz [1959] et Sharpe [1964], les modèles d'évaluation d'actifs financiers s'attachent à expliquer la rémunération du risque des actifs financiers en termes de rendements espérés. Ces modèles tendent à démontrer que la perception élémentaire du risque d'un revenu aléatoire à partir de sa variance est insuffisante. Dans les modèles de type CAPM (« Capital Asset Pricing Model »), on montre au contraire que c'est *la covariance avec certaines variables directrices* (plutôt que la variance) qui explique la différence entre rendements espérés et qui s'interprète donc comme la quantité de risque rémunéré. Ces variables directrices sont souvent appelées *facteurs* en tant que variables explicatives des rendements.

Dans un premier temps (section 7.2), on présentera cette notion traditionnelle de facteur en Finance sans détailler le cadre statistique. On utilisera pour cette synthèse le modèle de type CAPM le plus moderne introduit par Epstein et Zin [1989]. Dans un second temps (section 7.3), on cherchera à mettre en évidence les résultats de la synergie apparue dans les années quatre vingt entre Finance et Statistique :

- D'un côté, les statisticiens ont fourni aux financiers une modélisation de type ARCH qui a donné un contenu empirique à leurs modèles d'évaluation faisant intervenir les moments conditionnels (à l'information disponible sur le marché) de vecteurs aléatoires de grande taille (rendements des différents actifs, facteurs exogènes, ...).
- En sens inverse, les statisticiens ont bénéficié des idées de la Finance, en particulier pour permettre des paramétrisations plus parcimonieuses des modèles GARCH multivariés. Comme on l'a vu dans le chapitre 5, on sait en effet que dans un cadre multivarié, les modèles GARCH non contraints introduisent beaucoup trop des paramètres inconnus. Les modèles d'évaluation d'actifs financiers ont, en particulier par les facteurs qu'ils ont proposés, permis de justifier des modèles dynamiques plus contraints.

Le but de ce chapitre est précisément de dresser un catalogue de modèles statistiques dynamiques qui peuvent trouver certains fondements structurels dans des modèles d'évaluation d'actifs financiers. On se restreindra pour simplifier aux modèles à facteurs en temps discret.

Une illustration des applications possibles de l'approche mise en place dans ce chapitre sera détaillée dans la section 7.4 à propos de l'évaluation de la performance de portefeuilles.

7.2 LA NOTION DE FACTEUR EN FINANCE

7.2.1 Le problème de l'actualisation de revenus risqués

i) L'aversion par rapport au risque

Considérons un individu qui à une date t donnée cherche à évaluer un actif financier. Cet actif générera à la date $t + 1$ un gain γ_{t+1} , qui apparaît aléatoire à la date t . Une évaluation naïve de cet actif conduirait à l'apprécier en fonction du gain espéré $E_t \gamma_{t+1}$, la notation E_t signifiant que l'espérance est calculée conditionnellement à l'information I_t disponible à la date t . Cette appréciation est naïve dans la mesure où elle ne prend pas en compte la *risque* attaché à l'actif. En particulier, si l'individu considéré a de l'aversion pour le risque (hypothèse assez réaliste), il choisira entre deux actifs apportant le même gain espéré, celui qui est le « moins risqué », selon son critère subjectif d'appréciation du risque. Ce critère est usuellement défini en économie à partir d'une fonction d'utilité dite de Von Neumann-Morgenstern parce que ces deux mathématiciens lui ont donné des fondements axiomatiques (caractérisant la rationalité des choix dans l'incertain). Cette fonction d'utilité u , caractéristique des préférences de l'individu vis-à-vis de l'incertitude, est une fonction numérique de variable réelle, croissante au sens large et concave.

On suppose que l'individu préfère les actifs qui lui apportent la plus grande espérance de l'utilité du gain $E_t u(\gamma_{t+1})$. La concavité de u assure, en particulier par l'inégalité de Jensen, que :

$$E_t u(\gamma_{t+1}) \leq u(E_t \gamma_{t+1})$$

ce qui signifie que, compte tenu de son aversion au risque, l'individu préférerait se voir promettre pour la date $t + 1$ la richesse certaine (dès la date t) $E_t \gamma_{t+1}$ plutôt que la richesse aléatoire de même espérance γ_{t+1} .

ii) Actifs et portefeuilles

On suppose l'existence d'un nombre fini d'actifs indexés par $i, i = 1, \dots, n$. L'acquisition d'une unité de l'actif i à la date t revient à p_{it} , $i = 1, \dots, n$ et $p_t = (p_{1t}, \dots, p_{nt})'$ désigne le vecteur des coûts d'acquisition unitaires.

Nous désignons par $\gamma_{t+1} = (\gamma_{1t+1}, \dots, \gamma_{nt+1})'$ le vecteur des gains unitaires à la date $t + 1$ associés aux actifs détenus sur la période $(t, t + 1)$. Le rendement brut de l'actif i sur la période est $R_{it+1} = \gamma_{it+1}/p_{it}$.

Un portefeuille composé à la date t est décrit par un vecteur $\alpha_t = (\alpha_{1t}, \dots, \alpha_{nt})'$ donnant les quantités α_{it} des divers actifs dont il est composé. Un tel portefeuille a un coût d'acquisition à la date t $\alpha_t' p_t$ et génère à la date $t + 1$ un gain $\alpha_t' \gamma_{t+1}$; son rendement brut apparaît aléatoire à la date t et vaut $R_{pt+1}(\alpha_t) = \alpha_t' \gamma_{t+1} / \alpha_t' p_t$. Ce rendement s'exprime en fonction des rendements des divers actifs :

$$R_{pt+1}(\alpha_t) = \frac{\alpha_t' \gamma_{t+1}}{\alpha_t' p_t} = \sum_{i=1}^n \vartheta_{it} R_{it+1} \tag{7.1}$$

où $\vartheta_{it} = \alpha_{it} p_{it} / \alpha_t' p_t$. Il s'agit d'une moyenne de ces rendements pondérés par les coûts relatifs d'acquisition des divers actifs ; en particulier le rendement du portefeuille α_t ne dépend des quantités α_{it} que par l'intermédiaire des coûts relatifs ϑ_{it} .

iii) Choix de portefeuille optimal sur une période

Considérons un individu souhaitant acquérir à la date t des actifs pour un coût global W_{0t} . Il peut a priori retenir tout portefeuille α_t tel que $\alpha_t' p_t = W_{0t}$. Pour sélectionner un portefeuille parmi ceux ayant ce coût, on suppose qu'il maximise une utilité espérée $E_t u(\alpha_t' \gamma_{t+1})$. Les conditions du premier ordre de cette maximisation sous contrainte impliquent immédiatement (si u peut être dérivée sous l'opérateur espérance) que les prix $p_{it}, i = 1, \dots, n$, soient proportionnels aux quantités $E_t [\gamma_{it+1} u'(\alpha_t' \gamma_{t+1})]$. On appellera *facteur d'escompte* δ le coefficient de proportionalité tel que :

$$p_{it} = \delta E_t [\gamma_{it+1} u'(\alpha_t' \gamma_{t+1})] \quad \forall i = 1, \dots, n. \tag{7.2}$$

On remarquera que le gain aléatoire γ_{t+1} n'est pas actualisé uniquement par un facteur fixe δ , mais aussi par une fonction $u'(\alpha'_t \gamma_{t+1}) = u'[W_{0t} R_{pt+1}(\alpha_t)]$ du rendement aléatoire $R_{pt+1}(\alpha_t)$ du portefeuille α_t .

iv) Aspect intertemporel des programmes de consommation-investissement

Le modèle d'optimisation considéré ci-dessus est naïf dans la mesure où l'individu cherche à maximiser l'utilité espérée de sa richesse en fin de période ($t, t + 1$) sans s'inquiéter de son avenir au delà de la date $t + 1$; il s'agit d'un comportement myope. Un individu qui envisagerait plus rationnellement un programme de consommations présente et futures, $C_t, C_{t+1}, C_{t+2}, \dots$ devrait en effet choisir son portefeuille de façon à être indifférent à la marge entre un investissement supplémentaire dans un actif i ou la consommation immédiate.

L'utilité marginale $u'(C_t)$ représente le gain d'utilité apporté par la consommation à la date t d'une unité de bien supplémentaire (en confondant variation infinitésimale et variation d'une unité). L'unité de mesure des prix étant ici l'unité de bien de consommation, on peut donc interpréter le produit $p_{it}u'(C_t)$ comme le supplément d'utilité procuré à la date t par la consommation de la richesse p_{it} . Inversement, si cette richesse avait été investie dans une unité d'actif i , elle aurait procuré à la date $t + 1$ (en cas de consommation) un supplément d'utilité égal à $\gamma_{t+1}u'(C_{t+1})$. L'optimisation intertemporelle de l'utilité espérée conduit ainsi (voir Lucas [1978]) à écrire l'égalité en espérance (à un facteur d'escompte δ près) de ces deux gains d'utilité marginaux :

$$p_{it}u'(C_t) = \delta E_t [\gamma_{t+1}u'(C_{t+1})] \quad \forall i = 1, \dots, n. \quad (7.3)$$

Les équations (7.3) sont connues dans la littérature sous le nom d'équations d'Euler de maximisation de l'utilité espérée intertemporelle :

$$\sum_{h=0}^{\infty} \delta^h E_t u(C_{t+h}) \quad (7.4)$$

On remarquera l'analogie formelle entre (7.3) et (7.2) ; il s'agit dans les deux cas d'une formule d'actualisation, le facteur d'actualisation comportant une partie aléatoire dont la variable directrice est le rendement du portefeuille dans le premier cas, et la consommation dans le second. On voit ainsi apparaître des facteurs explicatifs des rendements dont le rôle va être mis en exergue par les modèles d'évaluation décrits ci-dessous.

v) Un modèle d'évaluation synthétique : le modèle d'Epstein et Zin

Les deux facteurs essentiels des modèles d'évaluation modernes sont le rendement du portefeuille $R_{mt+1} = R_{pt+1}(\alpha_t)$, et la consommation C_{t+1} d'un individu dit *représentatif* parce qu'il est censé représenter à lui seul l'ensemble des investisseurs. Autrement dit R_{mt+1} est le rendement d'un portefeuille α_t qui serait constitué avec l'ensemble des actifs en circulation sur le marché ; R_{mt+1} est

appelé pour cette raison *rendement du marché*, et est souvent assimilé en première approximation avec un indice représentatif de l'évolution de la capitalisation boursière. De la même façon C_{t+1} est la consommation en volume de toute la population considérée ; la croissance de cette consommation peut être assimilée (si on néglige le problème des biens durables) à la croissance du produit national brut.

Dans un article récent Epstein et Zin [1989] proposent la formule d'évaluation suivante :

$$p_{it} = \delta E_t \left[\gamma_{it+1} \left(\frac{C_{t+1}}{C_t} \right)^{\psi(\rho-1)} R_{mt+1}^{\psi-1} \right] \quad \forall i = 1, \dots, n \quad (7.5)$$

où δ , ψ et ρ sont trois paramètres réels donnés. Il ne s'agit pas ici de discuter la modélisation sophistiquée des choix dans l'incertain intertemporel qui a conduit Epstein et Zin à cette formule. On remarquera seulement que deux spécifications très particulières de cette formule s'interprètent bien à partir des modèles de choix de portefeuille décrits dans les paragraphes précédents.

D'une part si ψ est égal à 1, on retrouve exactement l'équation (7.3), appliquée à un individu représentatif dont la fonction d'utilité s'écrirait $u(C_t) = C_t^\rho / \rho$. Ce modèle d'évaluation, qui ne fait intervenir que la consommation (l'effet du facteur rendement du marché a été anihilé par la contrainte $\psi = 1$), est dû à Lucas [1978] et est désigné dans la littérature par les initiales CCAPM (« Consumption-based CAPM »).

Inversement, si $\rho = 1$, on retrouve exactement l'équation (7.2) (le facteur d'escompte étant maintenant noté $\delta \cdot W_{0t}^{1-\psi}$ au lieu de δ) appliquée à un individu représentatif dont la fonction d'utilité s'écrirait $u(\alpha'_t \gamma_{t+1}) = (\alpha'_t \gamma_{t+1})^\psi / \psi$. Ce modèle d'évaluation, qui ne fait intervenir que le rendement du marché (l'effet du facteur consommation a été anihilé par la contrainte $\rho = 1$), est le plus simple puisqu'il néglige l'aspect intertemporel du choix de portefeuille. Il s'agit en fait d'une version généralisée du CAPM traditionnel de Sharpe [1964], Lintner [1965], et Mossin [1966].

Depuis l'apparition de ce modèle simple jusqu'à la généralisation (7.5) qu'en ont donné Epstein et Zin [1989], les chercheurs ont été confrontés au paradoxe suivant : le CCAPM, a priori plus satisfaisant pour l'économiste puisque prenant en compte la dimension intertemporelle, expliquait moins bien les prix observés des actifs que le CAPM traditionnel. Le modèle d'Epstein et Zin reconcilie les deux approches en emboîtant les deux modèles dans un modèle plus général ; de plus, on peut montrer que les paramètres ρ et ψ de ce modèle général ont des interprétations économiques satisfaisantes (en termes d'élasticité de substitution intertemporelle et de mesure d'aversion au risque) même en dehors des deux cas extrêmes où l'un des deux est contraint à être égal à 1.

On peut remarquer en outre que l'emboîtement proposé par Epstein et Zin consiste à faire une moyenne géométrique (avec la pondération ψ) des facteurs

d'actualisation considérés respectivement par le CAPM et le CCAPM dans deux cas particuliers : CAPM avec utilité logarithmique et CCAPM avec utilité C_t^ρ / ρ . Si on confond moyenne géométrique et moyenne arithmétique, il s'agit en fait du procédé général d'emboîtement de modèles statistiques proposé par Davidson et MacKinnon [1981].

7.2.2 Les modèles à facteurs associés au modèle d'Epstein et Zin

i) La notion de relation multibêta

Les modèles d'évaluation d'actifs financiers travaillent traditionnellement en termes de rendements plutôt qu'en termes de gains ; ainsi l'équation d'Epstein et Zin (7.5) s'écrira dorénavant :

$$1 = \delta E_t \left[R_{it+1} \left(\frac{C_{t+1}}{C_t} \right)^{\psi(\rho-1)} R_{mt+1}^{\psi-1} \right] \quad \forall i = 1, \dots, n. \quad (7.6)$$

Supposons, pour simplifier l'interprétation à ce stade, qu'il existe en outre un actif non risqué, numéroté 0, dont le rendement R_{0t+1} à la date $t + 1$ est connu dès la date t . Si cet actif peut être incorporé dans le choix de portefeuille, il vérifie lui aussi une équation du type (7.6), si bien qu'on peut écrire par différence :

$$0 = E_t \left[(R_{it+1} - R_{0t+1}) \left(\frac{C_{t+1}}{C_t} \right)^{\psi(\rho-1)} R_{mt+1}^{\psi-1} \right] \quad \forall i = 1, \dots, n. \quad (7.7)$$

Cette équation explique les rendements nets espérés $E_t R_{it+1} - R_{0t+1}$, $i = 1, \dots, n$, par une mesure du risque des divers actifs i fondée sur la covariance conditionnelle à l'information disponible à la date t (notée Cov_t) entre leur rendement et la variable directrice $Z_{t+1} = (C_{t+1}/C_t)^{\psi(\rho-1)} R_{mt+1}^{\psi-1}$:

$$E_t R_{it+1} - R_{0t+1} = - \frac{\text{Cov}_t(R_{it+1}, Z_{t+1})}{E_t Z_{t+1}} \quad \forall i = 1, \dots, n \quad (7.8)$$

On obtient bien ainsi un modèle d'évaluation d'actifs financiers à un facteur Z_{t+1} au sens où le risque rémunéré de l'actif i est proportionnel à la covariance entre son rendement et le facteur. Ce risque est évidemment aussi proportionnel au coefficient bêta cher aux financiers, c'est-à-dire au coefficient de régression affine de R_{it+1} par rapport à Z_{t+1} :

$$\beta [R_{it+1}/Z_{t+1}] = \frac{\text{Cov}_t(R_{it+1}, Z_{t+1})}{\text{Var}_t Z_{t+1}}. \quad (7.9)$$

On parlera plus généralement de *relations d'évaluation multibêtas* pour caractériser des modèles écrivant les rendements nets espérés des divers actifs

comme combinaisons linéaires de plusieurs coefficients bêtas par rapport à plusieurs facteurs. En dehors des deux cas extrêmes du CAPM et du CCAPM on a plutôt l'habitude cependant d'écrire le modèle d'Epstein et Zin comme un modèle à deux facteurs : consommation et rendement du marché. Cela vient du fait que ces deux effets sont mélangés à l'intérieur de la variable directrice Z_{t+1} . Pour décomposer ces deux effets de façon linéaire (dans l'esprit des relations multibêtas) on s'intéressera au cas usuel en finance où les rendements sont supposés lognormaux (conditionnellement à l'information disponible en début de période). Rappelons que cette hypothèse de normalité *conditionnelle*, motivée par l'additivité des logarithmes des rendements sur des périodes consécutives, n'est pas contradictoire avec l'aspect leptokurtique des séries financières.

ii) Relation d'Epstein et Zin sous l'hypothèse de lognormalité

La linéarisation des conditions (7.6) est aisée si l'on suppose que la loi jointe du vecteur de taille $(n + 2)$:

$$\begin{aligned}
 Y_{t+1} &= \left[\log \frac{C_{t+1}}{C_t}, \log R_{mt+1}, \log R_{1t+1}, \dots, \log R_{nt+1} \right]' \\
 &= [Y_{ct+1}, Y_{mt+1}, Y_{1t+1}, \dots, Y_{nt+1}]'
 \end{aligned}$$

sachant l'information I_t disponible à la date t est gaussienne. Cette hypothèse est usuelle dans le cadre du CCAPM (voir par exemple Hansen et Singleton [1983]) puisque le logarithme d'un rendement ou d'un taux de croissance de la consommation sont additifs sur des périodes consécutives. Toutefois la présence du rendement du marché dans les composantes de Y_{t+1} rend cette hypothèse quelque peu paradoxale, puisque le rendement d'un portefeuille est une combinaison linéaire des rendements des actifs qui le composent. On admet cependant cette approximation en considérant que le portefeuille du marché contient d'autres actifs que les actifs de référence $i = 1, \dots, n$. On notera dans la suite :

$$\left\{ \begin{aligned}
 m_{it} &= E_t \log R_{it+1}, \sigma_{it}^2 \\
 &= \text{Var}_t \log R_{it+1}, \quad i = 1, \dots, n \\
 m_{mt} &= E_t \log R_{mt+1}, \sigma_{mt}^2 \\
 &= \text{Var}_t \log R_{mt+1} \\
 m_{ct} &= E_t \log(C_{t+1}/C_t), \sigma_{ct}^2 \\
 &= \text{Var}_t \log(C_{t+1}/C_t) \\
 \sigma_{ijt} &= \text{Cov}_t(\log R_{it+1}, \log R_{jt+1}) \\
 \sigma_{ict} &= \text{Cov}_t[\log R_{it+1}, \log(C_{t+1}/C_t)], \quad i, j \in \{1, 2, \dots, n, m\}
 \end{aligned} \right. \tag{7.10}$$

La relation (7.6) s'écrit alors :

$$\begin{aligned}
 0 = & -\log \delta + m_{it} + \psi(\rho - 1)m_{ct} + (\psi - 1)m_{mt} \\
 & + \frac{1}{2} \{ \sigma_{it}^2 + \psi^2(\rho - 1)^2 \sigma_{ct}^2 + (\psi - 1)^2 \sigma_{mt}^2 \} \\
 & + \psi(\rho - 1)\sigma_{ict} + (\psi - 1)\sigma_{imt} \\
 & + \psi(\rho - 1)(\psi - 1)\sigma_{mct} \quad i = 1, \dots, n
 \end{aligned} \tag{7.11}$$

Pour simplifier l'interprétation de (7.11) il est utile, comme on l'a déjà fait, d'introduire un actif non risqué, numéroté 0, dont le rendement R_{0t+1} vérifierait aussi (7.11), avec $m_{0t} = \log R_{0t+1}$, connu dès la date t (en particulier $\sigma_{0t}^2 = \sigma_{0ct} = \sigma_{0mt} = 0$). On obtient alors par différence :

$$E_t \log R_{it+1} - \log R_{0t+1} = \psi(1 - \rho)\sigma_{ict} + (1 - \psi)\sigma_{imt} - \frac{\sigma_{it}^2}{2}. \tag{7.12}$$

On obtient donc comme annoncé un modèle à deux facteurs, la prime de risque de l'actif i étant déterminée par une combinaison linéaire de ses covariances conditionnelles avec la consommation et le marché. Cette interprétation usuelle est cependant un peu abusive puisque la variance conditionnelle de $\log R_{it+1}$ apparaît aussi dans la formule d'évaluation (7.12).

Notons que le même modèle à facteurs peut aussi être obtenu *sans utiliser* l'hypothèse de lognormalité à partir d'un développement limité de (7.6) en supposant que $\log R_{it+1} \# (R_{it+1} - 1)$, pour $i = 1, 2, \dots, n, m$, et que de même $\log(C_{t+1}/C_t)$ est peu différent du taux de croissance de la consommation.

iii) Interprétations alternatives des facteurs

Une condition d'orthogonalité du type (7.6) ou (7.7) peut, selon les besoins de l'interprétation, être traduite de multiples façons en termes de modèle à 1, 2, 3, ... facteurs. Des interprétations à 1 puis à 2 facteurs ont déjà été données respectivement par (7.8) et (7.12). Une façon usuelle d'obtenir 3 facteurs consiste à introduire l'aléa de l'inflation. Les formules d'évaluation (7.2), (7.3) et (7.5) ne s'interprètent bien en effet en termes de comportement d'un individu représentatif que si les quantités qu'elles mettent en jeu sont exprimées en francs constants. Supposons alors que le taux de l'inflation sur la période $(t, t + 1)$ soit π_{t+1} , si bien que les rendements nominaux R_{it+1}^* et la consommation en valeur C_{t+1}^* sont reliés à leurs équivalents en francs constants de la date t par les relations :

$$\begin{cases} R_{it+1}^* = R_{it+1}\pi_{t+1}, & i = 1, 2, \dots, n, m \\ C_{t+1}^* = C_{t+1}\pi_{t+1}. \end{cases} \tag{7.13}$$

On déduit alors de (7.6) que :

$$1 = \delta E_t \left[\frac{1}{\pi_{t+1}^{\psi\rho}} R_{it+1}^* \left(\frac{C_{t+1}^*}{C_t} \right)^{\psi(\rho-1)} (R_{mt+1}^*)^{\psi-1} \right] \quad \forall i = 1, \dots, n. \quad (7.14)$$

Il est alors clair qu'avec un raisonnement analogue à celui qui a permis d'obtenir (7.12), on pourrait déduire de (7.14) un modèle à trois facteurs : consommation, inflation et marché, qui pourrait prendre une forme simple sous l'hypothèse qu'il existe un actif numéroté 0 dont le rendement *nominal* est sans risque.

Notons enfin que le facteur π_{t+1} introduit dans (7.14) pourrait avoir d'autres interprétations : taux de change, marge actuarielle, . . . Ces différents taux pourraient même intervenir simultanément. Il est cependant important de remarquer que ces facteurs ne sont pas nécessairement tous de même nature : certains sont observables, d'autres non (la consommation instantanée d'un individu représentatif n'est pas par exemple toujours observable), certains sont *évaluables*, d'autres non, au sens où ils représentent ou non le rendement d'un actif évalué par le marché. Par exemple, le facteur « rendement du marché » est évaluable, au sens où on peut écrire une relation d'évaluation (7.6) pour $i = m$. Le facteur consommation ne sera pas en général évaluable, à moins d'imaginer un titre indéré sur le produit national brut.

En anticipant sur la section statistique 7.3, on imagine bien que seuls poseront problème les facteurs non évaluables non observables. Une classification des modèles à facteurs sera construite dans cet esprit dans la section 7.3 ; on va déjà montrer ici sur l'exemple de la consommation comment un facteur non évaluable non observable peut être éliminé des relations d'évaluation.

iv) Modèles à facteurs ne faisant pas intervenir la consommation

Une raison importante de la popularité persistante du CAPM de Sharpe-Lintner-Mossin par rapport à des modèles intertemporels plus sophistiqués est l'aversion du praticien de la finance pour des modèles faisant intervenir des variables économiques abstraites difficilement observables comme la consommation. De plus on peut se demander si les mauvaises performances empiriques de modèles comme le CCAPM ne sont pas simplement dues à des erreurs de mesure sur la consommation. Ce problème est si prégnant que le simple titre « Intertemporal Asset Pricing Without Consumption Data » a pu récemment être un bon argument de vente d'un article (voir Campbell [1992]). Nous proposons dans cette section un travail d'élimination de la consommation analogue à celui de Campbell, mais fondé sur une hypothèse différente.

Alors que Campbell suppose que les relations non linéaires de composition des taux peuvent être linéarisées, nous nous contenterons pour notre part de remplacer la consommation par un facteur de risque de taux d'intérêt endogène. On supposera pour cela que l'information I_t disponible à la date t est engendrée

non seulement par l'observation de Y_τ , $\tau \leq t$, mais aussi d'une part par l'observation du taux sans risque $m_{0\tau} = \log R_{0\tau+1}$, $\tau \leq t$, et d'autre part de K variables exogènes X_τ , $\tau \leq t$.

On considèrera en fait dans la suite deux processus de dimension $n+2+K$ supposés stationnaires au second ordre : le processus

$$\tilde{Y}_t = [Y_t', X_t']',$$

et le processus

$$Z_t = [\log R_{0t+1}, \log R_{mt}, \log R_{1t}, \dots, \log R_{nt}, X_t']'.$$

Le modèle statistique paramétrique usuel dans ce contexte (voir par exemple Hansen et Singleton [1983], Evans et Wachtel [1990],...) consiste à supposer que, conditionnellement à I_t , le vecteur Y_{t+1} est gaussien de variance régulière $\Sigma(I_t)$ et d'espérance conditionnelle donnée sous une forme autoregressive. Il est par ailleurs usuel en finance (en particulier pour la spécification de modèles en temps continu) de considérer qu'il est toujours possible de se donner un ensemble suffisamment important de variables d'état pour que les processus multivariés considérés en temps discret soient markoviens d'ordre 1. C'est pourquoi on s'intéressera dans cette section aux variables d'état X_t qui assurent que \tilde{Y}_t soit markovien d'ordre 1. On se placera pour simplifier dans cette section d'interprétation financière dans un contexte d'homoscédasticité conditionnelle ($\Sigma(I_t) = \Sigma$ matrice constante) où les variables d'état X_t sont « exogènes » par rapport à la consommation au sens où, relativement à l'information I_t le processus de consommation ne cause pas le processus X :

$$\begin{aligned} E(X_{t+1} | I_t) &= E \left\{ X_{t+1} \mid [(Y_{it})'_{i=m,1,2,\dots,n}, X_t']'_{\tau \leq t} \right\} \\ &= b_X + \Phi_X' [(Y_{it})'_{i=m,1,2,\dots,n}, X_t']' \end{aligned} \quad (7.15)$$

où b_X et Φ_X sont des matrices données, la dernière égalité provenant du caractère markovien gaussien des processus considérés. En revanche le travail d'élimination de la consommation que nous allons entreprendre n'a évidemment de sens que parce que celle-ci cause le vecteur des rendements :

$$\begin{aligned} E[(Y_{it+1})_{i=m,1,2,\dots,n} | I_t] \\ \neq E \left\{ (Y_{it+1})_{i=m,1,2,\dots,n} \mid [(Y_{it})'_{i=m,1,2,\dots,n}, X_t']'_{\tau \leq t} \right\}. \end{aligned} \quad (7.16)$$

On se propose ici de montrer que dans la mesure où le modèle financier (7.6) n'est pas remis en question, le modèle statistique peut très généralement être formulé de manière équivalente pour le processus Z_t , ce qui présente l'avantage

de permettre une étude statistique ne portant que sur des données financières (sans utiliser les données de consommation peu fiables). On montrera en même temps que le processus d'innovation de $\log(C_{t+1}/C_t)$ est une forme linéaire du processus d'innovation de Z_t , ce qui permet de faire apparaître (7.12) comme une relation multibêta n'utilisant pas la consommation.

On part du modèle général (7.12) de la prime de risque, écrit non seulement pour les actifs élémentaires $i = 1, \dots, n$, mais aussi pour le rendement du marché :

$$m_{it} = m_{0t} + \psi(1 - \rho)\sigma_{ict} + (1 - \psi)\sigma_{imt} \tag{7.17}$$

$$- \frac{\sigma_{it}^2}{2} \quad \forall i = 1, 2, \dots, n, m.$$

Ce modèle est complété par la relation d'évaluation de l'actif sans risque :

$$m_{0t} = \log \delta + \psi(1 - \rho)m_{ct} + (1 - \psi)m_{mt} \tag{7.18}$$

$$+ \frac{1}{2} \left\{ \psi^2(\rho - 1)^2\sigma_{ct}^2 + (\psi - 1)^2\sigma_{mt}^2 + 2\psi(\rho - 1)(\psi - 1)\sigma_{mct} \right\}.$$

Pour que cette relation exprime aussi m_{ct} en fonction de m_{0t} , on élimine dans (7.18) m_{mt} grâce à (7.17) écrit pour $i = m$:

$$m_{ct} = (1 - \rho)^{-1} \left\{ m_{0t} + \log \delta + \frac{1}{2} \left[\psi(\rho - 1)^2\sigma_{ct}^2 + (1 - \psi)\sigma_{mt}^2 \right] \right\}. \tag{7.19}$$

Le modèle financier produit donc les équations (7.17) et (7.19) exprimant $E_t Y_{t+1}$ en fonction de m_{0t} et des moments du second ordre.

Sous l'hypothèse d'homoscédasticité conditionnelle, le modèle financier montre donc en particulier que la dynamique des rendements est complètement déterminée par la dynamique soit du rendement de l'actif sans risque (voir (7.17)) soit par celle de la consommation (voir (7.19)). Cette remarque est à la base du théorème ci-dessous, qui montre qu'au lieu de définir la dynamique à partir de celle du processus de consommation⁽¹⁾ on peut tout aussi bien la définir à partir de celle du processus de taux d'intérêt sans risque, étant entendu que

$$m_{0t+1} = \log R_{0t+2}$$

apparaît risqué à la date t , puisque ce taux sans risque sur la période $(t + 1, t + 2)$ ne sera en général connu qu'à la date $t + 1$. Alors que

$$Y_{it+1} = \log R_{it+1}, i = 1, 2, \dots, n, m$$

⁽¹⁾ Voir par exemple Giovannini et Jorion [1989] et Tauchen et Hussey [1991], qui se donnent pour la consommation une dynamique AR(1) avec résidus ARCH.

s'interprète comme le taux de rendement d'un placement dans l'actif i sur la période $(t, t + 1)$ (en confondant comme d'habitude $\log R_{it+1}$ et $R_{it+1} - 1$),

$$Y_{0t+1} = \log R_{0t+2}$$

ne s'interprète comme le taux de rendement d'un placement sur la période $(t, t + 1)$ qu'à condition d'imaginer une obligation à taux variable indexée sur le taux sans risque à venir.

Théorème 7.1

Sous les hypothèses (7.15), (7.16), (7.17) et (7.19), les deux propriétés suivantes sont équivalentes :

(P1) \tilde{Y}_{t+1} est conditionnellement gaussien sachant I_t , de variance régulière constante Σ avec :

$$E(Y_{ct+1}|I_t) = b_c + \Phi'_c \tilde{Y}_t,$$

où $Y_{ct+1} = \log(C_{t+1}/C_t)$.

(P2) Z_{t+1} est conditionnellement gaussien sachant I_t , de variance régulière constante Ω avec :

$$E(Y_{0t+1}|I_t) = d_0 + \Pi'_0 Z_t$$

et Y_{0t} est une fonction affine de \tilde{Y}_t .

De plus, si ces propriétés sont vérifiées, \tilde{Y}_{t+1} et Z_{t+1} sont des processus VAR(1) gaussiens (pour un choix convenable d'une condition initiale), le processus d'innovation de \tilde{Y}_{t+1} (resp. Z_{t+1}) étant une fonction linéaire du processus d'innovation de Z_{t+1} (resp. \tilde{Y}_{t+1}).

Ce théorème, démontré en annexe A, prouve que le modèle statistique usuel relatif au processus \tilde{Y}_t peut être formulé de manière équivalente pour le processus Z_t . Il n'y a donc aucune nécessité d'utiliser des données de consommation, une variable de taux d'intérêt pouvant toujours remplacer la variable d'état « taux de croissance de la consommation ». Ainsi, on a montré, comme dans le chapitre 5, que des facteurs exogènes peuvent toujours être remplacés par des facteurs endogènes ; bien entendu, les arguments financiers utilisés ici pour cette substitution ne sont pas exactement équivalents aux arguments statistiques invoqués dans le chapitre 5. Ce type de substitution est aussi largement pratiqué dans les modèles de la structure par terme des taux d'intérêt⁽²⁾ où l'on peut remplacer des variables d'état représentatives de la technologie, de l'inflation, . . . par des taux d'intérêt à des échéances particulières (taux long, taux court, . . .). On pourrait d'ailleurs montrer⁽³⁾ que des taux de rendement d'obligations de maturité plus

⁽²⁾ Voir par exemple Cox, Ingersoll et Ross [1985a] et [1985b].

⁽³⁾ Voir Pastorello et Renault [1992].

longue peuvent être utilisés de manière analogue au théorème 7.1 ci-dessus pour éliminer la consommation dans le contexte de processus autoregressifs d'ordre supérieur à 1.

La liaison linéaire entre les innovations des processus \tilde{Y} et Z permet en outre d'écrire que :

$$Y_{ct+1} - E_t Y_{ct+1} = \sum_{i=0}^n \alpha_i (Y_{it+1} - E_t Y_{it+1}) \tag{7.20}$$

$$+ \alpha_m (Y_{mt+1} - E_t Y_{mt+1}) + \alpha'_X (X_{t+1} - E_t X_{t+1})$$

où $\alpha_0, \alpha_1, \dots, \alpha_n, \alpha_m$ et les coefficients de α_X sont des réels donnés ($\alpha_0 \neq 0$). Par conséquent pour tout actif j donné, σ_{jct} est aussi la covariance conditionnelle à I_t entre $\log R_{jt+1}$ et la variable

$$F_{t+1} = \sum_{i=0}^n \alpha_i Y_{it+1} + \alpha_m Y_{mt+1} + \alpha'_X X_{t+1}.$$

Autrement dit, la relation multibêta (7.12) peut aussi être écrite :

$$E_t \log R_{jt+1} - \log R_{0t+1}$$

$$= \psi(1 - \rho) \text{Cov}_t(\log R_{jt+1}, F_{t+1}) + (1 - \psi)\sigma_{imt} - \frac{\sigma_{it}^2}{2}.$$

Il s'agit encore d'un modèle financier à deux facteurs, mais ceux-ci peuvent maintenant s'interpréter comme les rendements de deux actifs financiers : le portefeuille du marché et un portefeuille qui combinerait les actifs de rendement Y_{it+1} , $i = 0, 1, 2, \dots, n, m$, en proportion α_i , $i = 0, 1, 2, \dots, n, m$ et de rendement X_{kt+1} , $k = 1, \dots, K$, en proportions définies par α_X . Cette interprétation suppose bien sûr que l'on admette qu'il existe K actifs dont les rendements sont parfaitement corrélés avec les variables d'état X_{kt+1} , $k = 1, \dots, K$; c'est la notion de « mimicking factor » usuelle en finance⁽⁴⁾. En tout cas, la consommation a effectivement été remplacée par le « mimicking factor » F_{t+1} . Comme nous l'avons dit plus haut, Campbell [1992] a obtenu plusieurs interprétations analogues, mais au prix de linéarisations approchées de la contrainte de budget intertemporelle.

7.3 CLASSIFICATION DES MODÈLES A FACTEURS

Le premier principe de discrimination entre les facteurs sera leur position vis-à-vis du modèle financier : facteurs endogènes (c'est-à-dire évalués par le modèle financier) ou facteurs exogènes (variables de conjoncture déterminées

⁽⁴⁾ Voir Huberman, Kandel et Stambaugh [1987].

à l'extérieur de la sphère financière). Cette opposition endogène/exogène ne présuppose pas le caractère endogène ou non au sens statistique du terme⁽⁵⁾ des variables considérées. Au contraire, la classification statistique endogène/exogène viendra compléter la distinction liée au modèle financier.

On se place dans un modèle financier qui aboutit à une relation d'évaluation du type de celles considérées dans la section 7.2 :

$$\lambda = E_t \left[R_{it+1} \cdot \prod_{j=1}^{\ell} X_{jt+1} \cdot \prod_{h=1}^m \tilde{R}_{ht+1} \right] \quad (7.21)$$

pour les actifs risqués $i = 1, 2, \dots, n$. Dans cette relation λ est une constante réelle donnée, les variables X_{jt+1} , $j = 1, \dots, \ell$, représentent des facteurs exogènes au sens financier (par exemple le taux de croissance de la consommation), alors que les variables \tilde{R}_{ht+1} , $h = 1, \dots, m$, sont des facteurs dits endogènes parce que la relation (7.21) est valide en particulier pour $R_{it+1} = \tilde{R}_{ht+1}$; les facteurs endogènes, comme par exemple le rendement du marché, sont eux-mêmes évalués par le modèle financier. Un facteur comme le taux d'inflation (voir subsection (iii) de la section 7.2.2, pp. 142–143) peut être endogène ou exogène selon qu'il existe ou non une possibilité de dupliquer le taux d'inflation par le rendement d'un portefeuille financier. De façon générale, on peut imaginer un modèle financier faisant intervenir simultanément un grand nombre de facteurs endogènes et/ou exogènes : consommation, production, inflation, taux de change, rendements des différents marchés, ... On considèrera d'abord dans la section 7.3.1 des contextes où les facteurs exogènes sont directement reliés à des facteurs statistiquement observables.

7.3.1 Le rôle des facteurs exogènes

i) Modèles dynamiques à facteurs

Les modèles dynamiques à facteurs, définis dans la section 5.2.2, peuvent trouver un fondement structurel dans une relation d'évaluation du type (7.21). Le rôle des facteurs F_t fondant le modèle statistique (5.6) est alors celui d'un résumé statistique qui permet d'oublier les caractéristiques individuelles de l'agent représentatif intervenant éventuellement dans (7.21) (fonction d'utilité, consommation, ...). On se permettra pour cela un léger changement de notations par rapport à la section précédente, en posant

$$Y_{t+1} = [R_{it+1} - R_{0t+1}]_{1 \leq i \leq n},$$

vecteur des rendements nets des n actifs considérés. Une relation d'évaluation générale du type (7.21) peut toujours alors être mise par différence sous la forme :

$$E_t(Y_{it+1} \cdot Z_{t+1}) = 0 \quad \forall i = 1, 2, \dots, n \quad (7.22)$$

⁽⁵⁾ Voir Gourieroux et Monfort [1989].

où la variable directrice Z_{t+1} (variable aléatoire réelle) est un agrégat qui contient les facteurs explicatifs du modèle financier, et en particulier les caractéristiques individuelles de l'agent représentatif (voir par exemple (7.7) et (7.8)). Un vecteur

$$F_{t+1} = [F_{kt+1}]_{1 \leq k \leq K},$$

de K facteurs observables, peut alors être introduit pour résumer l'effet de l'agrégat (en général non observable) Z_{t+1} sur la dynamique des rendements. On suppose plus précisément que :

$$Z_{t+1} = b_{0t} + \sum_{k=1}^K b_{kt} F_{kt+1}, \quad k = 1, \dots, K \quad (7.23a)$$

$$E_t(Y_{it+1}|F_{t+1}) = \mu_{it} + \sum_{k=1}^K \beta_{ik} F_{kt+1} \quad (7.23b)$$

$$\text{Var}_t(Y_{t+1}|F_{t+1}) = \Omega \quad (7.23c)$$

où les réels μ_{it} , β_{ik} , b_{kt} , ainsi que les coefficients la matrice Ω sont déterministes (compte tenu de l'information disponible à la date t). On remarquera que les « coefficients bêtas » de la régression de Y_{it+1} sur les différents facteurs, de même que le « risque idiosyncratique » Ω (la variabilité des rendements non expliquée par les facteurs) sont supposés invariants dans le temps. En effet, si ces coefficients dépendaient de l'information disponible à la date t , les facteurs F_{kt+1} , $k = 1, 2, \dots, K$, ne joueraient pas complètement leur rôle de *résumé* de l'information pertinente pour prévoir les gains ou les rendements. Il faut remarquer à ce sujet que l'équation (7.22) ne contraint Y_{it+1} qu'à une homothétie près (de rapport appartenant à l'information I_t) si bien que (7.23) pourra dans la suite selon le contexte être interprétée en considérant soit que Y_{it+1} est un taux de rendement net ($Y_{it+1} = R_{it+1} - R_{0t+1}$), soit que Y_{it+1} est un gain net ou « valeur actuelle nette » ($Y_{it+1} = \gamma_{it+1} - R_{0t+1} p_{it}$). En outre, les coefficients b_{kt} de la combinaison affine qui relie la variable directrice Z_{t+1} aux facteurs F_{kt+1} peuvent eux dépendre de l'information disponible à la date t .

Puisque d'après (7.23) :

$$\mu_{it} = E_t Y_{it+1} - \sum_{k=1}^K \beta_{ik} E_t F_{kt+1}$$

on voit que les facteurs F_{kt+1} sont définis à une translation déterministe près (compte tenu de l'information disponible à la date t), une telle translation ne faisant que modifier la valeur des intercepts μ_{it} et b_{0t} . On peut en particulier toujours choisir une translation de façon que :

$$E_t(F_{kt+1} \cdot Z_{t+1}) = 0, \quad \forall k = 1, 2, \dots, K. \quad (7.24)$$

Il suffit pour cela d'imposer que :

$$E_t F_{kt+1} = - \frac{\text{Cov}_t(F_{kt+1}, Z_{t+1})}{E_t Z_{t+1}}$$

en admettant que $E_t Z_{t+1}$ est non nul, ce qui est le cas dans tous les modèles usuels. Par exemple, dans (7.2) ou (7.3), $Z_{t+1} = u'(\alpha'_t \gamma_{t+1})$ ou $Z_{t+1} = u'(C_{t+1})$, avec u fonction strictement croissante et donc Z_{t+1} est à valeurs strictement positives. La relation d'évaluation (7.22), qui s'écrit :

$$E_t Y_{it+1} \cdot E_t Z_{t+1} + \text{Cov}_t(Y_{it+1}, Z_{t+1}) = 0$$

peut alors, compte tenu de (7.23) et (7.24), être écrite sous les formes équivalentes :

$$E_t Y_{it+1} \cdot E_t Z_{t+1} + \text{Cov}_t [E_t(Y_{it+1} | F_{t+1}), Z_{t+1}] = 0$$

d'après (7.23a), ou :

$$E_t Y_{it+1} \cdot E_t Z_{t+1} + \sum_{k=1}^K \beta_{ik} \text{Cov}_t(F_{kt+1}, Z_{t+1}) = 0$$

d'après (7.23b) ou encore :

$$E_t Y_{it+1} = \sum_{k=1}^K \beta_{ik} E_t F_{kt+1}$$

d'après (7.24). Remarquons enfin que (7.24) est automatiquement vérifiée si les facteurs sont endogènes, puisque si

$$F_{kt+1} = \sum_{i=1}^n a_{ki} Y_{it+1},$$

(7.24) se déduit immédiatement de (7.22).

On a donc démontré :

Théorème 7.2

En postulant le modèle à facteurs (7.23), ceux-ci peuvent toujours être choisis de façon à vérifier aussi (7.24), et alors la relation d'évaluation (7.22) est caractérisée par la nullité de l'intercept μ_{it} . De plus, (7.24) est impliquée par (7.22) dans le cas de facteurs endogènes.

Ainsi, la relation d'évaluation (7.22) se caractérise (sous l'hypothèse (7.23)) par un modèle dynamique à facteurs au sens de (5.6) :

$$\begin{cases} E_t(Y_{t+1} | F_{t+1}) = B F_{t+1} \\ \text{Var}_t(Y_{t+1} | F_{t+1}) = \Omega. \end{cases} \quad (7.25)$$

Gourieroux, Monfort et Renault [1991b] ont montré comment, dans le contexte d'un modèle de ce type, il est possible d'identifier statistiquement le sous-ensemble des facteurs pertinents, ou au contraire de tester l'oubli de facteurs. De tels tests (en particulier tests sur le noyau, l'image et le rang de la matrice B) permettent de comparer la validité des différents modèles financiers proposés, en donnant une réponse statistique à des questions du type : la consommation est-elle un facteur essentiel ?, le rendement du marché mesuré par tel indice boursier est-il un résumé convenable ?, . . . L'application de ces tests à l'évaluation de la performance de portefeuilles sera discutée dans la section 7.4.

ii) Endogénéisation statistique des facteurs exogènes

Le modèle dynamique à facteurs considéré dans la section précédente ne s'intéresse qu'à la régression d'un vecteur Y_{t+1} de taux de rendements sur un vecteur F_{t+1} de facteurs conditionnellement à l'information I_t engendrée par l'observation jusqu'à la date t des rendements et des facteurs F_τ , $\tau \leq t$. La loi marginale des facteurs F_t n'est jamais modélisée.

La tradition initiée par Hansen et Singleton [1983] dans un contexte d'homoscédasticité conditionnelle et perpétuée par Giovannini et Jorion [1989], Tauchen et Hussey [1991], . . . dans un contexte ARCH consiste au contraire à se donner la loi marginale des facteurs F_t pour en déduire la loi jointe des rendements et des facteurs. On a déjà montré dans la section 7.2 comment la donnée de la dynamique de la consommation (ou de manière équivalente du taux sans risque) pouvait, dans le contexte du modèle d'Epstein et Zin, déterminer la dynamique de tous les rendements. Considérons plus généralement le modèle (7.21) en admettant que la loi jointe du logarithme de toutes les variables considérées conditionnelle à l'information I_t soit normale. Les équations (7.21) permettent alors (de manière analogue à ce qui a été fait dans (7.11)) d'exprimer les prévisions optimales $E_t \log R_{it+1}$ des rendements en fonction des prévisions optimales des autres variables et des variances-covariances conditionnelles. Plusieurs approches statistiques sont alors envisageables :

- Soit comme Hansen et Singleton [1983] on suppose que le vecteur des logarithmes des variables considérées suit un processus vectoriel autorégressif gaussien (et donc homoscédastique). Les équations (7.21) apparaissent alors comme un système de contraintes sur les paramètres de ce processus VAR (contraintes qui définissent en outre éventuellement des paramètres auxiliaires introduits dans (7.21) comme δ , ρ et ψ dans le cas particulier (7.6)). Les paramètres d'intérêt peuvent alors être estimés de manière asymptotiquement efficace par le maximum de vraisemblance contraint. Dans la mesure où tous les facteurs sont observables, la seule difficulté est numérique parce que les contraintes mélangent les paramètres d'espérance et de variance (voir par exemple (7.11)), ce qui interdit d'estimer simplement par moindres carrés. Si de plus certains facteurs comme

la consommation posent des problèmes d'observabilité, les paramètres d'intérêt peuvent ne pas être tous identifiables. On a proposé par exemple dans notre Théorème 1 de remplacer le processus VAR consommation-rendements par un processus VAR taux d'intérêt-rendements ; ce faisant on perd partiellement l'identifiabilité des paramètres individuels ρ , δ et ψ (voir Pastorello et Renault [1992]).

- L'autre approche consiste à se donner simplement la loi du processus des variables d'état, et à en déduire par le modèle d'évaluation (7.21) la loi du processus des rendements. Ceci permet en particulier d'aboutir à une spécification de l'hétéroscédasticité conditionnelle des rendements par construction compatible avec le modèle d'évaluation. C'est ainsi que Giovannini et Jorion [1989], dans le cadre du modèle d'Epstein et Zin [1989] spécifient un modèle autoregressif d'ordre 1 pour la consommation et supposent que le vecteur des innovations (consommation et rendements) suit un processus GARCH(1,1) diagonal (voir chapitre 5). Ils peuvent alors, à partir de (7.11) et de l'hypothèse de lognormalité jointe, inférer la loi jointe du processus consommation-rendements. Bonomo et Garcia [1991] et Tauchen et Hussey [1991] ont une approche plus structurelle en recherchant les paramètres de la fonction d'utilité qui rendent compatibles les caractéristiques des rendements observés avec l'équation d'évaluation (7.3) du CCAPM. Ils ne postulent pas la lognormalité des rendements, et considèrent un modèle très général de la dynamique de la consommation ; ce modèle très général peut toujours être approché par un modèle à facteurs discrets (voir section 5.3) si bien qu'il est possible de chercher directement par « calibrage » les valeurs des paramètres d'intérêt qui rendent le mieux compte des évidences empiriques.

7.3.2 Le rôle des facteurs endogènes

Le caractère endogène (vis-à-vis du modèle financier) de certains facteurs peut permettre soit, si nécessaire, d'éliminer des facteurs qui posent un problème d'observabilité statistique, soit, sinon, de réduire la dimension de l'incertitude naturellement associée au modèle financier.

i) Élimination de facteurs non observables

Les facteurs non observables explicatifs de la dynamique du prix d'un actif i intéressent le financier à deux niveaux : celui du rendement espéré $E_t R_{it+1}$, et celui du risque associé à ce rendement, mesuré par la variance conditionnelle $\text{Var}_t R_{it+1}$ et aussi éventuellement (en particulier dans une optique de couverture) par la covariance conditionnelle $\text{Cov}_t(R_{it+1}, R_{jt+1})$ avec le rendement d'autres actifs. Il s'agit donc de généraliser les résultats de la subsection (iv) de la section 7.2.2, subsection où l'on a montré comment il était possible d'éliminer le facteur consommation dans un contexte d'homoscédasticité conditionnelle. Cette généralisation peut être envisagée dans deux directions que l'on étudiera

séparément pour simplifier :

- D'une part, on peut s'intéresser à un modèle financier du type (7.21) plus général que celui d'Epstein et Zin, sans remettre en question l'hypothèse d'homoscédasticité conditionnelle.
- D'autre part, on peut au contraire considérer un modèle statistique plus général, autorisant l'hétéroscédasticité conditionnelle.

a) Modèle général avec homoscédasticité

Si les logarithmes des variables intervenant dans la relation d'évaluation (7.21) constituent de manière jointe un processus vectoriel autoregressif d'ordre p , qui engendre la filtration d'information I_t (on pourrait aisément généraliser comme dans le Théorème 1 au cas où le caractère autoregressif n'est obtenu qu'après ajout d'autres variables exogènes), on déduit de (7.21) (avec des notations évidentes) que :

$$E_t \log R_{it+1} + \sum_{j=1}^{\ell} E_t \log X_{jt+1} + \sum_{h=1}^m E_t \log \tilde{R}_{ht+1} = a_i \quad (7.26)$$

où a_i est une constante réelle donnée. (7.26) définit un système de contraintes sur les paramètres d'espérance du processus VAR \tilde{Y}_t avec :

$$\tilde{Y}_t = \left[\log R_{it+1}, i = 1, \dots, n; \log X_{jt+1}, j = 1, \dots, \ell; \log \tilde{R}_{ht+1}, h = 1, \dots, m \right]'$$

Si ces paramètres ne peuvent être estimés par le maximum de vraisemblance contraint parce que certaines composantes de \tilde{Y}_t (certains facteurs exogènes $\log X_{jt+1}$) sont non observables, on peut généraliser le Théorème 1 en fondant l'inférence statistique sur un processus VAR Z_t observable parce que les composantes $\log X_{jt+1}$ ont été remplacées par des facteurs évalués par le marché. De tels facteurs sont bijectivement reliés aux facteurs exogènes par l'équation (7.26), dont on déduit aisément l'analogie de la preuve du Théorème 1 à partir de (7.17).

b) Modèle d'Epstein et Zin avec hétéroscédasticité

.Le principe de démonstration rappelé ci-dessus ne fonctionne plus en présence d'hétéroscédasticité conditionnelle ; en effet dans ce cas le coefficient a_i introduit dans (7.26) dépend lui même de l'information disponible à la date t , si bien que l'observation du seul rendement de l'actif i associé à la contrainte (7.26) ne permet pas de caractériser à la fois les contraintes sur la dynamique des moments du premier et du second ordre. Deux types de solution peuvent à ce stade être envisagés bien qu'ils aient encore peu été développés à ce jour dans la littérature :

- Puisque l'on ne peut pas espérer déduire d'une seule contrainte (7.26) une spécification des deux dynamiques des moments du premier et du second ordre, on peut chercher à caractériser les restrictions imposées par (7.26) à une dynamique générale. Il s'agirait donc de concevoir une notion de modèle ARCH-M multivarié dans lequel espérances et variances covariances conditionnelles évoluerait conjointement. On manque encore à ce jour de modèles parcimonieux satisfaisant pour ce type de problème.
- L'autre solution consiste à ajouter à (7.26) une autre contrainte fournie par l'observation du prix d'un actif *dérivé* de l'actif i (actif dont le prix sera directement relié à celui de l'actif i). C'est ce prix qui jouera le rôle du facteur endogène qui nous manque. De ce point de vue, les prix d'options sont particulièrement intéressants parce que une option d'achat (ou de vente) de l'actif i a un prix directement fonction de la *volatilité* (écart type conditionnel du rendement) de celui-ci.

Considérons, à titre d'illustration de cette dernière solution, la méthode proposée par Renault et Touzi [1992], dans le contexte d'un modèle à volatilité stochastique introduit par Hull et White [1987]. L'idée de ce modèle est de considérer le processus de prix S_t d'une action sous la forme :

$$dS_t = \mu S_t dt + \sigma_t S_t dB_{1t} \quad (7.27)$$

où B_{1t} est un mouvement brownien standard et σ_t est un processus indépendant de B_{1t} . On peut envisager en particulier le cas où une certaine fonction $f(\sigma_t)$ suit un processus d'Ornstein-Uhlenbeck :

$$df(\sigma_t) = a [b - f(\sigma_t)] dt + cd B_{2t} \quad (7.28)$$

avec B_{2t} mouvement brownien standard indépendant de B_{1t} . Un tel modèle a déjà été considéré au chapitre 6 (voir (6.25)) avec $f(\sigma_t) = \log \sigma_t$, mais B_{2t} éventuellement corrélé avec B_{1t} . L'hypothèse de non corrélation est relativement compatible avec l'observation statistique et simplifiera beaucoup les raisonnements ci-dessous puisqu'elle permettra de caractériser indépendamment les dynamiques du prix de l'option et du prix de l'action support. De façon générale $f(\cdot)$ doit être une bijection connue de R_+^* dans R , puisque σ_t s'interprète comme la volatilité du processus de prix :

$$\sigma_t^2 = \lim_{h \rightarrow 0} \frac{1}{h} \text{Var}_t \left[\frac{S_{t+h}}{S_t} \right] \quad (7.29)$$

σ_t^2 caractérise donc de manière instantanée la variance conditionnelle du rendement d'un investissement dans l'action de prix S_t . L'espérance conditionnelle de celui-ci est caractérisée par le terme constant μ :

$$\mu = \lim_{h \rightarrow 0} \frac{1}{h} E_t \left[\frac{S_{t+h}}{S_t} - 1 \right]. \quad (7.30)$$

Le modèle d'Hull et White (7.27) est la généralisation naturelle du célèbre modèle de Black et Scholes [1973] dans lequel la volatilité σ_t était supposée constante. Le résultat de ces modèles, conçus pour l'évaluation d'options, peut être retrouvé simplement à partir du modèle d'Epstein et Zin. Une option d'achat émise à la date t sur l'action considérée et pouvant être exercée à la date $t+1$ procurera à son détenteur à cette date le gain $\gamma_{t+1} = \max\{0, S_{t+1} - K\}$, où K est le prix d'exercice de cette option. Son prix à la date t doit donc selon le modèle d'Epstein et Zin être donné par (voir (7.5)) :

$$P_t = \delta E_t \left[\max(0, S_{t+1} - K) \left(\frac{C_{t+1}}{C_t} \right)^{\psi(\rho-1)} R_{mt+1}^{\psi-1} \right]. \quad (7.31)$$

On a par ailleurs bien entendu toujours la relation d'Epstein et Zin pour la valorisation de l'action :

$$S_t = \delta E_t \left[S_{t+1} \left(\frac{C_{t+1}}{C_t} \right)^{\psi(\rho-1)} R_{mt+1}^{\psi-1} \right]. \quad (7.32)$$

Ce sont ces deux équations qui vont pouvoir être utilisées simultanément avec les équations de valorisation du rendement du marché et de l'actif sans risque pour éliminer non seulement le facteur exogène consommation, mais aussi le facteur exogène qui est introduit dans le processus de volatilité (en temps continu ce facteur est représenté par le brownien B_{2t}). Sans donner les détails techniques⁽⁶⁾, on peut déjà remarquer que pour le calcul des espérances (7.31) et (7.32) l'indépendance des browniens B_{1t} et B_{2t} assure que, conditionnellement à l'information I_t (engendrée ici par $B_{1\tau}$ et $B_{2\tau}$, $\tau \leq t$), et conditionnellement à la valeur du processus de volatilité σ_τ pour $t \leq \tau \leq t+1$, $\log(S_{t+1}/S_t)$ suit la loi normale de moyenne $\mu - V_t/2$, et de variance V_t , avec :

$$V_t = \int_t^{t+1} \sigma_\tau^2 d\tau. \quad (7.33)$$

Conditionnellement à la trajectoire de la volatilité, on récupère donc dans ce modèle la propriété de lognormalité des rendements, et on remarque de plus que cette loi conditionnelle ne dépend de la trajectoire de la volatilité que par l'intermédiaire de la valeur moyenne de celle-ci sur la période $(t, t+1)$. Il est donc relativement naturel de prolonger cette propriété en supposant que, conditionnellement à I_t et à la valeur de V_t , la loi jointe du processus $[\log(S_{t+1}/S_t), \log(C_{t+1}/C_t), \log R_{mt+1}]'$ est normale. Sous cette hypothèse on peut montrer (voir Renault et Touzi [1992]) que :

$$P_t = S_t \left\{ E_t \left[\Phi \left(\frac{x_t}{\sqrt{V_t}} + \frac{\sqrt{V_t}}{2} \right) \right] - e^{-x_t} E_t \left[\Phi \left(\frac{x_t}{\sqrt{V_t}} - \frac{\sqrt{V_t}}{2} \right) \right] \right\} \quad (7.34)$$

⁽⁶⁾ Nous renvoyons le lecteur à Renault et Touzi [1992].

où Φ est la fonction de repartition de la loi normale centrée et réduite, et $x_t = \log(S_t R_{0t+1}/K)$. Dans la démonstration de (7.34) à partir de (7.31), la consommation, les paramètres individuels ρ , ψ et δ , ainsi que le taux de rendement instantané espéré μ ont été éliminés grâce aux relations de valorisation de l'actif sans risque, du portefeuille du marché et de l'action. P_t , calculé par (7.34), est en fait une fonction déterministe de σ_t (par l'intermédiaire du conditionnement implicite dans l'opérateur E_t), et l'on montre dans Renault et Touzi [1992] que cette fonction est bijective. Le prix de l'option peut donc bien remplacer pour l'inférence statistique l'observation du facteur exogène de volatilité; cette remarque est utilisée systématiquement dans Renault et Touzi [1992] pour estimer les paramètres du processus d'Ornstein-Uhlenbeck (7.28). La problématique est donc très différente de celle envisagée au chapitre 6, où l'estimation de ces paramètres a été étudiée sans utiliser des observations sur le prix d'options.

ii) La réduction de la dimension de l'incertitude

Il a été montré au chapitre 5 (voir section 5.2.3) que la notion de facteur endogène permettait de définir des paramétrages parcimonieux de modèles GARCH multivariés. Ces paramétrages étaient obtenus en considérant K facteurs $F_t = a'Y_t$, à coefficients fixes sur les rendements (ou les gains) nets Y_{it} (un facteur est défini par une colonne de la matrice a) qui constituaient eux mêmes des processus GARCH. On réduisait ainsi substantiellement (dans la mesure où K est petit devant la dimension n de Y_t) la dimension de l'espace décrit par la trajectoire de $\text{Var}_t Y_{t+1}$.

On se propose dans cette section de relier davantage cette hypothèse à la théorie financière. On se placera pour simplifier dans le cadre du modèle (7.2) de choix de portefeuille optimal sur une période. En notant ici $Y_{it+1} = \gamma_{it+1} - R_{0t+1} p_{it}$ le gain net de l'actif i , on déduit de la relation d'évaluation (7.2) écrite à la fois pour l'actif i et pour l'actif sans risque :

$$p_{it} = \delta E_t[\gamma_{it+1} u'(\alpha'_t \gamma_{t+1})] \quad \forall i = 1, \dots, n.$$

et

$$1 = \delta E_t [R_{0t+1} u'(\alpha'_t \gamma_{t+1})]$$

que par différence :

$$0 = E_t [Y_{it+1} u'(G_{t+1})] \quad \forall i = 1, \dots, n \quad (7.35)$$

où $G_{t+1} = \alpha'_t \gamma_{t+1}$ désigne le gain du portefeuille optimal α_t . La relation fondamentale (7.35) va nous permettre d'introduire successivement la notion de portefeuille efficient et celle de portefeuilles de base définissant les facteurs exogènes.

a) La notion de portefeuille efficient

Supposons dans un premier temps que le vecteur Y_{t+1} des gains nets (ou de manière équivalente le vecteur des rendements) soit gaussien conditionnellement à l'information I_t disponible à la date t . Notons que la lognormalité et la normalité sont sensiblement équivalentes dans la mesure où l'on confond comme c'est usuel $\log R_{it+1}$ avec le taux de rendement $R_{it+1} - 1$. L'avantage ici de la loi normale est sa stabilité par combinaison linéaire et la propriété de linéarité de l'espérance conditionnelle :

$$E_t(Y_{it+1}|G_{t+1}) = \mu_{it} + \beta_{it}G_{t+1} \tag{7.36}$$

puisque G_{t+1} est une fonction linéaire du vecteur gaussien Y_{t+1} (conditionnellement à I_t , μ_{it} et β_{it} sont déterministes).

La condition (7.35) :

$$0 = E_t Y_{it+1} \cdot E_t u'(G_{t+1}) + \text{Cov}_t[Y_{it+1}, u'(G_{t+1})]$$

s'écrit donc, compte tenu de (7.36) :

$$0 = E_t Y_{it+1} \cdot E_t u'(G_{t+1}) + \beta_{it} \text{Cov}_t[(G_{t+1}, u'(G_{t+1}))]. \tag{7.37}$$

On remarque que $\text{Cov}_t[G_{t+1}, u'(G_{t+1})]$ est négatif (u' est décroissante puisque u est concave) si bien que (7.37) traduit l'idée essentielle du CAPM de Sharpe, Lintner et Mossin : les gains nets espérés sont une fonction linéaire croissante des coefficients bêtas par rapport au portefeuille optimal. C'est la notion de *droite de marché*, cas particulier des relations multibêtas décrites dans la section 7.2. On retiendra à ce stade que l'individu choisit un portefeuille G_{t+1} tel que :

$$E_t Y_{t+1} = \xi_t \text{Cov}_t(Y_{t+1}, G_{t+1})$$

où ξ_t est un réel connu à la date t (on rappelle que le coefficient bêta est défini par $\beta_{it} = \text{Cov}_t(Y_{it+1}, G_{t+1}) / \text{Var}_t(G_{t+1})$). Puisque $G_{t+1} = \alpha'_t Y_{t+1}$, on trouve :

$$\alpha_t = \frac{1}{\xi_t} [\text{Var}_t Y_{t+1}]^{-1} E_t Y_{t+1}. \tag{7.38}$$

Ainsi, à un facteur multiplicatif près, le portefeuille optimal α_t ne dépend pas des caractéristiques individuelles de l'agent résumées dans la fonction d'utilité $u(\cdot)$. La notion de portefeuille de marché censé être détenu par un agent «représentatif», prend donc ici tout son sens puisque tous les portefeuilles individuels sont proportionnels au portefeuille de composition $\bar{x}_t = [\text{Var}_t Y_{t+1}]^{-1} E_t Y_{t+1}$. Le portefeuille de marché, somme des portefeuilles individuels, sera donc lui même proportionnel à \bar{x}_t , et tous ces portefeuilles auront le même rendement R_{mt+1} .

On peut considérer à titre d'illustration le cas particulier d'une fonction d'utilité exponentielle : $u(G_{t+1}) = -\exp\{-\eta G_{t+1}\}$, où η est une constante

positive mesurant l'aversion pour le risque de l'individu ; η est généralement appelé « *indice d'Arrow-Pratt d'aversion absolue pour le risque* »⁽⁷⁾. Dans ce cas, sous l'hypothèse de normalité conditionnelle des rendements :

$$E_t u(G_{t+1}) = -\exp\left(-\eta E_t G_{t+1} + \frac{\eta^2}{2} \text{Var}_t G_{t+1}\right).$$

On voit donc ici que la maximisation de l'utilité espérée se ramène à la minimisation de la variance du gain, pour un gain espéré donné. On peut ainsi montrer que l'ensemble des portefeuilles proportionnels à \bar{x}_t , c'est-à-dire l'ensemble des portefeuilles susceptibles d'être optimaux pour une certaine fonction d'utilité $u(\cdot)$ dans un contexte gaussien, coïncide avec l'ensemble des portefeuilles dits *efficients* (au sens de Markowitz [1959]) parce qu'ils minimisent une certaine mesure du risque (la variance) à gain espéré donné.

b) Les portefeuilles de base

Les relations (7.35) et (7.36) peuvent dans certains cas permettre d'écrire un modèle dynamique à 1 facteur de la même façon qu'on a introduit des modèles dynamiques à K facteurs à partir de (7.22) et (7.23). Mais il faudrait pour cela que le coefficient β_{it} de la régression affine du gain net de l'actif i , Y_{it+1} , sur le gain net G_{t+1} du portefeuille ne dépende pas de l'information conditionnante I_t . Si tel n'est pas le cas, on peut imaginer que l'analogie des relations (7.23) soit vérifiée avec $Z_{t+1} = G_{t+1}$. Ceci est particulièrement intéressant pour l'application pratique si les facteurs F_{kt+1} , $k = 1, 2, \dots, K$, s'interprètent comme les gains nets de K portefeuilles de base de composition fixe :

$$F_{t+1} = a' Y_{t+1}. \quad (7.39)$$

Dans (7.39) les coefficients réels a_{ki} , $i = 1, \dots, n$, désignent les quantités des actifs $i = 1, \dots, n$ qui constituent le portefeuille de base numéroté k , si bien que le gain net F_{kt+1} de ce portefeuille s'écrit bien comme la combinaison linéaire :

$$F_{kt+1} = \sum_{i=1}^n a_{ki} Y_{it+1}.$$

Supposer, comme dans (7.23a), que G_{t+1} est une combinaison affine des F_{kt+1} revient donc à supposer qu'à chaque date le portefeuille optimal peut être constitué en combinant les K portefeuilles de base invariants dans le temps avec le placement sans risque ; même si les coefficients de cette combinaison affine peuvent, eux, varier, le nombre de transactions pour constituer à chaque date un portefeuille optimal est ainsi grandement minimisé si les portefeuilles de base sont proposés directement sur le marché (d'où l'importance des SICAV, ou des fonds communs de placement en pratique). Compte tenu de (7.38), cette

⁽⁷⁾ Voir par exemple Laffont [1983] pour l'interprétation de cet indice.

hypothèse est clairement caractérisée par la restriction suivante sur la dynamique des gains :

$$[\text{Var}_t Y_{t+1}]^{-1} E_t Y_{t+1} \in \text{Im}(a). \quad (7.40a)$$

Dans ce contexte la condition (7.23b), qui exprime la constance des coefficients bêtas de la matrice $\text{Cov}_t(Y_{t+1}, F_{t+1})[\text{Var}_t F_{t+1}]^{-1}$ de regression de Y_{t+1} sur F_{t+1} s'écrit :

$$(\text{Var}_t Y_{t+1}) a [a' (\text{Var}_t Y_{t+1}) a]^{-1} \quad \text{ne dépend pas de } I_t \quad (7.40b)$$

Enfin, la condition (7.23c) se caractérise évidemment par :

$$\left\{ \begin{array}{l} \text{Var}_t Y_{t+1} - (\text{Var}_t Y_{t+1}) a [a' (\text{Var}_t Y_{t+1}) a]^{-1} a' \text{Var}_t Y_{t+1} \\ \text{ne dépend pas de } I_t. \end{array} \right. \quad (7.40c)$$

On a ainsi démontré à partir du Théorème 7.2 un résultat démontré directement dans Gourieroux, Monfort et Renault [1991b] :

Théorème 7.3

Y_t a une représentation factorielle d'ordre r endogène avec des facteurs $F_t = a' Y_t$ si et seulement si la condition (7.40), conjonction des conditions (7.40a), (7.40b) et (7.40c) est vérifiée. La condition (7.40a) signifie que tout portefeuille efficient à la date t est défini par un vecteur de quantités investies α_t combinaison linéaire des colonnes de a .

On remarquera que le modèle à facteurs déduit du Théorème 7.3 :

$$\left\{ \begin{array}{l} Y_{t+1} = B F_{t+1} + u_{t+1}, \\ \text{Cov}_t(F_{t+1}, u_{t+1}) = 0, \\ \text{Var}_t u_{t+1} = \Omega \end{array} \right.$$

fait apparaître des coefficients bêtas et un risque idiosyncratique fixes quand ces quantités sont calculées en termes de gains, et non en termes de rendements, comme c'est le plus usuel⁽⁸⁾. Ces deux types d'hypothèses sont clairement en général incompatibles quand les prix relatifs des actifs varient. Un modèle analogue en termes de rendements aurait pû être justifié si l'on avait cherché des portefeuilles de base fixes non pas en termes de quantités investies (dans les différents actifs), mais en termes de coûts relatifs des investissements ; on a vu en effet (voir (7.1)) que les rendements se combinent selon les coûts relatifs. On est en fait plus intéressés par des portefeuilles de base fixes en termes de quantités investies à la fois pour une raison théorique, et pour une motivation pratique :

⁽⁸⁾ Voir par exemple Connor et Korajczyk [1989].

- La raison théorique est la référence au CAPM de Sharpe-Lintner-Mossin, dans lequel le portefeuille de référence est celui du marché. Pour une offre d'actifs approximativement invariante dans le temps, on préférera donc le point de vue des quantités fixes.
- La motivation pratique est celle de la mise à jour du portefeuille optimal à partir des portefeuilles de base ; ce que l'on peut obtenir sans transactions nouvelles est effectivement un portefeuille à quantités fixes dans le temps même si les coûts relatifs varient du fait de la variation des prix relatifs des différents actifs.

7.4 APPLICATION A L'EVALUATION DE PERFORMANCE DE PORTEFEUILLES

7.4.1 Indice de performance de Sharpe

On a montré dans la section 7.3.2.2 qu'un portefeuille α_t est efficient à la date t si et seulement si il existe un réel λ_t , tel que :

$$\alpha_t = \lambda_t \cdot [\text{Var}_t Y_{t+1}]^{-1} E_t Y_{t+1}$$

Un tel portefeuille procure un gain espéré :

$$E_t(\alpha'_t Y_{t+1}) = \lambda_t (E_t Y_{t+1})' [\text{Var}_t Y_{t+1}]^{-1} (E_t Y_{t+1})$$

au prix d'une variance :

$$\text{Var}_t(\alpha'_t Y_{t+1}) = \lambda_t^2 (E_t Y_{t+1})' [\text{Var}_t Y_{t+1}]^{-1} (E_t Y_{t+1}).$$

Autrement dit, le gain net espéré par unité de risque consenti (risque mesuré par l'écart type conditionnel de ce gain) est égal à :

$$\frac{E_t(\alpha'_t Y_{t+1})}{\sqrt{\text{Var}_t(\alpha'_t Y_{t+1})}} = \sqrt{(E_t Y_{t+1})' [\text{Var}_t Y_{t+1}]^{-1} (E_t Y_{t+1})}. \quad (7.41)$$

Le ratio « gain net espéré sur écart type du gain » d'un portefeuille a été appelé par Sharpe⁽⁹⁾ « *indice de performance* » du portefeuille ; la définition originelle a été donnée en termes de rendements, mais est clairement équivalente à une définition en termes de gains, puisque l'indice est sans unité. (7.41) montre que tous les portefeuilles efficients ont le même indice de performance ; cette valeur commune de la performance :

$$S_t = \sqrt{(E_t Y_{t+1})' [\text{Var}_t Y_{t+1}]^{-1} (E_t Y_{t+1})} \quad (7.42)$$

est clairement (par définition de l'efficacité moyenne- variance) égale à la valeur maximale des indices de performance de Sharpe des portefeuilles qui peuvent être

⁽⁹⁾ Voir aussi Dybvig et Ross [1985].

constitués à partir des actifs $i = 1, 2, \dots, n$ (et de l'actif sans risque). S_t est donc une caractéristique de l'ensemble des actifs $i = 1, 2, \dots, n$ à la date t , et a été appelé par Jobson et Korkie [1982] « *performance potentielle de l'ensemble des actifs $i = 1, 2, \dots, n$* ». Cependant, Jobson et Korkie [1982], se plaçant dans un contexte de rendements indépendants et identiquement distribués, n'ont pas fait la différence entre la *performance conditionnelle* S_t (performance conditionnelle à l'information disponible à la date t) et la *performance marginale* S :

$$S = \sqrt{(E Y_t)' [\text{Var } Y_t]^{-1} (E Y_t)} \tag{7.43}$$

définie en supposant (ce que l'on fera dans toute la suite) que le processus Y_t est stationnaire au second ordre ; cette hypothèse peut en pratique être plus ou moins justifiée selon que l'on l'applique au processus des gains (nets) ou des rendements (nets).

7.4.2 Performances conditionnelles et historiques

Une problématique usuelle en finance consiste à se demander si un sous-ensemble donné des actifs (K actifs parmi les n) est « efficient » au sens où un portefeuille efficient peut toujours être construit en utilisant uniquement ces K actifs (et l'actif sans risque). Quitte à renuméroter les actifs, on peut toujours supposer que les actifs d'intérêt sont les K premiers. On partitionnera donc le vecteurs des gains nets en :

$$Y_t = (Y'_{1t}, Y'_{2t})'$$

où Y_{1t} vecteur aléatoire de taille K représente les gains nets des K actifs candidats à l'efficience et Y_{2t} le vecteur des gains nets des $n - K$ actifs qu'on voudrait ne pas utiliser. Compte tenu de (7.40a), il s'agit donc de se demander si $[\text{Var}_t Y_{t+1}]^{-1} E_t Y_{t+1} \in \text{Im}(a)$, où a est la matrice qui définit les K « portefeuilles » d'intérêt à partir des n actifs, c'est-à-dire :

$$Y_{1t} = a' Y_t$$

ce qui correspond évidemment à :

$$a' = [\text{Id}_K, 0]. \tag{7.44}$$

Notons que toute l'étude qui suit pourrait aisément être généralisée à n'importe quelle matrice a de taille $n \times K$ de rang K ; il s'agirait alors de se demander si un ensemble donné de K portefeuilles définis par les colonnes de a est « efficient ». La notation adoptée ici signifie simplement que au moyen d'un changement de base, ces K portefeuilles ont été assimilés aux K premiers actifs.

Compte tenu de (7.44), la propriété d'efficience (7.40a) considérée pour les K premiers actifs se caractérise par le fait que les $n - K$ dernières lignes de

la matrice $[\text{Var}_t Y_{t+1}]^{-1} E_t Y_{t+1}$ sont nulles, ou encore que $E_t Y_{t+1}$ est une combinaison linéaire des K premières colonnes de $\text{Var}_t Y_{t+1}$, c'est-à-dire appartient à l'espace image de $\text{Cov}_t(Y_{t+1}, Y_{1t+1})$. Si on suppose de plus que la propriété (7.40b) est vérifiée pour a définie par (7.44), $\text{Cov}_t(Y_{t+1}, Y_{1t+1})[\text{Var}_t Y_{1t+1}]^{-1}$, ne dépend pas de t et s'écrit :

$$B_{\cdot 1} = \begin{bmatrix} \text{Id}_K \\ B_{21} \end{bmatrix}$$

où B_{21} est la matrice des coefficients bêtas (conditionnels à I_t) de Y_{2t+1} par rapport à Y_{1t+1} . Autrement dit, l'efficienne des K premiers actifs signifie que $E_t Y_{t+1}$ appartient à l'espace image de $B_{\cdot 1}$, ce qui se caractérise par :

$$E_t Y_{2t+1} = B_{21} E_t Y_{1t+1}. \quad (7.45)$$

Cette relation *multibêta conditionnelle* a souvent été considérée dans la littérature sous sa forme affaiblie (impliquée par (7.45)), forme dite *inconditionnelle*⁽¹⁰⁾ :

$$E Y_{2t+1} = B_{21} E Y_{1t+1}. \quad (7.46)$$

En d'autres termes, on se contentera, pour caractériser l'efficienne, de vérifier la nullité d'un intercept d_2 dans la régression conditionnelle de Y_{2t+1} sur Y_{1t+1} :

$$\begin{aligned} Y_{2t+1} &= d_2 + B_{21} Y_{1t+1} + u_{2t+1} & (7.47) \\ E_t u_{2t+1} &= 0 \\ \text{Cov}_t(u_{2t+1}, Y_{1t+1}) &= 0 \end{aligned}$$

au lieu de considérer une alternative plus générale à l'hypothèse d'efficienne dans laquelle d_2 serait une fonction de l'information disponible à la date t . Il faut cependant bien noter que l'on s'intéresse toujours à l'*efficienne conditionnelle* de l'ensemble des K premiers actifs (minimisation de la variance à espérance donnée dans la loi conditionnelle à I_t) et non à l'*efficienne inconditionnelle* (approche moyenne-variance dans la loi marginale). Il est néanmoins surprenant de constater que si l'on suppose que (7.47) est un modèle dynamique à facteurs au sens de (5.6) ($\text{Var}_t u_{2t+1} = \Omega_{22}$ indépendante de I_t), la constante $d'_t \Omega_{22}^{-1} d_2$ représente à la fois la différence des (carrés des) performances conditionnelles et marginales de Y_t et Y_{1t} . C'est l'objet du Théorème 7.4 suivant, dont nous empruntons la démonstration à Gourieroux, Monfort et Renault (1991a) (voir annexe B) :

⁽¹⁰⁾ Voir la discussion sur ce sujet de Gibbons et Ferson [1985].

Théorème 7.4

Si

$$Y_{2t+1} = d_2 + B_{21} Y_{1t+1} + u_{2t+1},$$

avec :

$$\begin{aligned} E_t u_{2t+1} &= 0, \\ \text{Cov}_t(u_{2t+1}, Y_{1t+1}) &= 0, \\ \text{Var}_t u_{2t+1} &= \Omega_{22}, \end{aligned}$$

alors

$$S_t^2 - S_{1t}^2 = S^2 - S_1^2 = d_2' \Omega_{22}^{-1} d_2$$

où :

$$\begin{aligned} S_{1t}^2 &= (E_t Y_{1t+1})' [\text{Var}_t Y_{1t+1}]^{-1} (E_t Y_{1t+1}) \\ S_t^2 &= (E_t Y_{t+1})' [\text{Var}_t Y_{t+1}]^{-1} (E_t Y_{t+1}) \\ S_1^2 &= (E Y_{1t})' [\text{Var} Y_{1t}]^{-1} (E Y_{1t}) \\ S^2 &= (E Y_t)' [\text{Var} Y_t]^{-1} (E Y_t). \end{aligned}$$

Ainsi, la nullité de d_2 caractérise à la fois l'efficacité conditionnelle et l'efficacité inconditionnelle de l'ensemble des K premiers actifs.

On peut donc se demander si le calcul usuel des performances « historiques », estimateurs naturels des performances marginales, ne suffirait pas pour le test statistique non seulement de l'efficacité inconditionnelle, mais aussi de l'efficacité conditionnelle, propriété beaucoup plus intéressante en pratique. Les performances historiques sont définies par :

$$\hat{S}^2 = (\hat{Y}_T)' [\hat{\Sigma}_T]^{-1} (\bar{Y}_T) \tag{7.48a}$$

$$\hat{S}_1^2 = (\bar{Y}_{1T})' [\hat{\Sigma}_{11T}]^{-1} (\bar{Y}_{1T}) \tag{7.48b}$$

où :

$$\bar{Y}_T = [\bar{Y}'_{1T}, \bar{Y}'_{2T}]' = T^{-1} \sum_{t=1}^T Y_t,$$

et :

$$\hat{\Sigma}_T = T^{-1} \sum_{t=1}^T (Y_t - \bar{Y}_T)(Y_t - \bar{Y}_T)' = \begin{bmatrix} \hat{\Sigma}_{11T} & \hat{\Sigma}_{12T} \\ \hat{\Sigma}_{21T} & \hat{\Sigma}_{22T} \end{bmatrix}.$$

7.4.3 Tests d'efficacité et de performance

Le test de l'efficacité des K premiers actifs a été étudié par Jobson et Korkie [1982] dans le contexte de rendements iid gaussiens. Dans ce cas il est

évidemment équivalent de tester l'efficacité conditionnelle ou inconditionnelle. Nous allons montrer ici que dans le cadre du modèle dynamique à facteurs (7.47) le test de l'efficacité conditionnelle prend exactement la même forme que celui proposé par Jobson et Korkie ; ce résultat, dû à Gourieroux, Monfort et Renault [1991a], généralise évidemment celui de Jobson et Korkie.

Il s'agit en fait de tester la nullité de l'intercept d_2 dans le modèle de regressions empilées :

$$Y_2 = e \cdot d_2' + Y_1 B_{21}' + U_2 \quad (7.49)$$

où Y_2 et Y_1 sont respectivement des matrices de taille $T \times (n - K)$ et $T \times K$, regroupant les observations Y_{2t} et Y_{1t} , $t = 1, \dots, T$, et e est le vecteur de taille T dont toutes les composantes valent 1. Comme Gourieroux, Monfort et Renault [1991b] l'ont montré, le test de la nullité de d_2 est fondé sur la somme des valeurs propres de la matrice :

$$R^2[Y_2, e|Y_1] = \Omega_{22T}^{-1} Y'(\text{Id} - P_1)e[e'(\text{Id} - P_1)e]^{-1} e'(\text{Id} - P_1)Y_2 \quad (7.50)$$

où P_1 désigne la matrice de projection orthogonale sur le sous-espace de R^T engendré par les colonnes de Y_1 . Cette matrice s'interprète bien en effet comme le carré d'une corrélation partielle entre Y_2 et e , conditionnellement à Y_1 .

Plus précisément, la statistique de test de Wald généralisé est asymptotiquement équivalente à :

$$\begin{aligned} \xi_T &= \text{Trace} \{ R^2[Y_2, e|Y_1] \} \\ &= [e'(\text{Id} - P_1)e]^{-1} \text{Trace} \{ \Omega_{22T}^{-1} Y_2'(\text{Id} - P_1)ee'(\text{Id} - P_1)Y_2 \} \\ &= [e'(\text{Id} - P_1)e]^{-1} [e'(\text{Id} - P_1)Y_2 \Omega_{22T}^{-1} Y_2'(\text{Id} - P_1)e]. \end{aligned}$$

Mais on remarque d'une part que :

$$\begin{aligned} e'(\text{Id} - P_1)e &= e'[\text{Id} - Y_1(Y_1'Y_1)^{-1}Y_1']e \\ &= T \left\{ 1 - Y_{1T}' \left(\hat{\Sigma}_{11T} + \bar{Y}_{1T}\bar{Y}_{1T}' \right)^{-1} \bar{Y}_{1T} \right\} \\ &= T \left\{ 1 - \bar{Y}_{1T}' \left(\hat{\Sigma}_{11T}^{-1} - (1 + \hat{\Sigma}_1^2)^{-1} \hat{\Sigma}_{11T}^{-1} \bar{Y}_{1T}\bar{Y}_{1T}' \hat{\Sigma}_{11T}^{-1} \right) \bar{Y}_{1T} \right\} \\ &= T \left\{ 1 - \hat{\Sigma}_1^2 + (1 + \hat{\Sigma}_1^2)^{-1} \hat{\Sigma}_1^4 \right\} \\ &= T \left(1 + \hat{\Sigma}_1^2 \right)^{-1}. \end{aligned}$$

De façon analogue, nous avons d'autre part :

$$Y_2'(\text{Id} - P_1)e = T(1 + \hat{\Sigma}_1^2)^{-1} \left[\bar{Y}_{2T} - \hat{\Sigma}_{21T} \hat{\Sigma}_{11T}^{-1} \bar{Y}_{1T} \right]. \quad (7.51)$$

Tenant compte du fait que :

$$\hat{\Omega}_{22T} = \hat{\Sigma}_{22T} - \hat{\Sigma}_{21T} \hat{\Sigma}_{11T}^{-1} \hat{\Sigma}_{12T},$$

nous en déduisons l'expression suivante de ξ_T :

$$T(1 + \hat{S}_1^2)^{-1} \left[\bar{Y}_{2T} - \hat{\Sigma}_{21T} \hat{\Sigma}_{11T}^{-1} \bar{Y}_{1T} \right]' \\ \times \left[\hat{\Sigma}_{22T} - \hat{\Sigma}_{21T} \hat{\Sigma}_{11T}^{-1} \hat{\Sigma}_{12T} \right]^{-1} \left[\bar{Y}_{2T} - \hat{\Sigma}_{21T} \hat{\Sigma}_{11T}^{-1} \bar{Y}_{1T} \right].$$

On trouve donc :

$$\xi_T = T(1 + \hat{S}_1^2)^{-1} (\hat{S}^2 - \hat{S}_1^2)$$

puisque, d'après le Théorème 7.4, $S^2 - S_1^2 = d_2' \Omega_{22}^{-1} d_2$, et $\bar{Y}_{2T} - \hat{\Sigma}_{21T} \hat{\Sigma}_{11T}^{-1} \bar{Y}_{1T}$ est bien l'estimateur de d_2 obtenu en remplaçant les moments théoriques par les moments empiriques. On a donc démontré :

Théorème 7.5

Dans le cadre du modèle à facteurs défini dans le Théorème 7.4, un équivalent asymptotique de la statistique de test de Wald de l'hypothèse $d_2 = 0$, c'est-à-dire de l'efficience conditionnelle de l'ensemble des K premiers actifs est définie par :

$$\xi_T = T \left(1 + \hat{S}_1^2 \right)^{-1} \left(\hat{S}^2 - \hat{S}_1^2 \right).$$

La loi asymptotique sous l'hypothèse nulle de cette statistique est une loi χ^2 à $n - K$ degrés de liberté.

L'expression de ξ_T est bien celle fournie par Jobson et Korkie [1982]. A un facteur multiplicatif près, la différence des carrés des performances historiques des deux ensembles d'actifs suffit donc bien à apprécier l'efficience conditionnelle du sous-ensemble des K premiers actifs.

ANNEXE A

Preuve du Théorème 7.1

Supposons la propriété (P1) vérifiée

On déduit immédiatement de (7.17) et (7.19) :

$$E(Y_{it+1}|I_t) = b_i + (1 - \rho)\Phi'_c \tilde{Y}_t,$$

où b_i est une constante réelle donnée pour $i = 1, 2, \dots, n, m$. Compte tenu de (7.15), on peut donc affirmer que :

$$E(\tilde{Y}_{t+1}|I_t) = b + \Phi \tilde{Y}_t,$$

où b est une matrice colonne donnée et Φ est une matrice constante définie par blocs par :

$$\Phi = \left[\begin{array}{c|c} \Phi'_c & \\ \hline (1 - \rho)\Phi'_c & \\ \vdots & \\ (1 - \rho)\Phi'_c & \\ \hline 0 & \Phi'_x \\ \vdots & \\ 0 & \end{array} \right].$$

Ceci montre bien comme annoncé que \tilde{Y}_{t+1} est un processus VAR(1) gaussien. On sait par ailleurs d'après (7.19) que :

$$E_t Y_{ct+1} = (1 - \rho)^{-1} Y_{0t} + a,$$

où a est une constante réelle donnée. En identifiant cette prévision avec celle fournie par le processus VAR ci-dessus, on en déduit comme annoncé que :

$$Y_{0t} = (1 - \rho)\Phi'_c \tilde{Y}_t + g,$$

où g est une constante réelle donnée. En particulier le processus d'innovation $Y_{0t+1} - E_t Y_{0t+1}$ est la forme linéaire :

$$(1 - \rho)\Phi'_c (\tilde{Y}_{t+1} - E_t \tilde{Y}_{t+1})$$

du processus d'innovation de \tilde{Y} .

Puisque d'après (7.19) :

$$\begin{aligned} Y_{0t+1} &= (1 - \rho) E_{t+1} Y_{ct+2} - (1 - \rho)a \\ &= (1 - \rho)\Phi'_c \tilde{Y}_{t+1} - (1 - \rho)a, \end{aligned}$$

on a :

$$E_t Y_{0t+1} = (1 - \rho)\Phi'_c E_t \tilde{Y}_{t+1} - (1 - \rho)a. \quad (A.1)$$

Compte tenu de la forme du processus VAR caractérisé ci-dessus, on sait que :

$$E_t \bar{Y}_{t+1} = [E_t Y_{ct+1}, (1 - \rho) E_t Y_{ct+1}, \dots, (1 - \rho) E_t Y_{ct+1}, E_t X'_{t+1}]'$$

(à une translation constante près), avec :

$$E_t Y_{ct+1} = (1 - \rho)^{-1} Y_{0t} + a$$

(d'après (7.19)) et :

$$E_t X_{t+1} = b_X + \Phi'_X [(Y_{it})'_{i=1,2,\dots,n,m}, X'_t]'$$

En reportant cette expression de $E_t \bar{Y}_{t+1}$ dans (A.1) on en déduit que :

$$E_t Y_{0t+1} = (1 - \rho) \Phi'_c \left[\begin{array}{c|ccc} (1 - \rho)^{-1} & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 0 \\ \hline 0 & & & & \\ \vdots & & & & \\ 0 & & & & \Phi'_X \end{array} \right] Z_t + f_0 \quad (A.2)$$

où f_0 est une constante réelle donnée. Puisque par ailleurs d'après (7.17) :

$$E_t Y_{it+1} = Y_{0t} + f_i,$$

f_i étant une constante réelle donnée, et puisque :

$$E_t X_{t+1} = b_X + \Phi'_X [(Y_{it})'_{i=1,2,\dots,n,m}, X'_t]'$$

on peut conclure que :

$$E_t Z_{t+1} = d + \Pi Z_t,$$

avec d vecteur colonne constant et Π matrice constante définie par blocs par :

$$\Pi = \left[\begin{array}{c|ccc} \Phi_{cc} & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \dots \\ 0 & 0 & 0 & \dots & 0 \end{array} \right] + \left[\begin{array}{c} (1 - \rho) \Phi_{c(c)} \\ \text{Id}_{n+K+1} \end{array} \right] \cdot \left[\begin{array}{c|ccc} 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 0 \\ \hline 0 & & & & \\ \vdots & & & & \\ 0 & & & & \Phi'_X \end{array} \right],$$

où $\Phi'_c = [\Phi_{cc}, \Phi_{c(c)}]$.

Pour obtenir la propriété (P2) il nous reste à vérifier que $\text{Var}_t Z_{t+1}$ est une matrice constante régulière. Mais, d'après la relation sur les innovations déjà obtenue :

$$Y_{0t+1} - E_t Y_{0t+1} = (1 - \rho) \Phi'_c (\bar{Y}_{t+1} - E_t \bar{Y}_{t+1}),$$

on a :

$$\text{Var}_t Z_{t+1} = \begin{bmatrix} (1-\rho)\Phi'_c \\ 0 \\ \vdots \\ \vdots \\ \text{Id}_{n+K+1} \\ 0 \end{bmatrix} \cdot \Sigma \cdot \begin{bmatrix} (1-\rho)\Phi'_c \\ 0 \\ \vdots \\ \vdots \\ \text{Id}_{n+K+1} \\ 0 \end{bmatrix}.$$

Cette matrice est régulière, car le premier coefficient de Φ'_c est non nul, d'après l'hypothèse (7.16). D'où (P2), le caractère VAR(1) gaussien du processus Z_t et la relation annoncée entre les innovations.

Supposons la propriété (P2) vérifiée

On déduit immédiatement de (7.17) :

$$E(Y_{it+1}|I_t) = d_i + Y_{0t},$$

où d_i est une constante réelle pour $i = 1, 2, \dots, n, m$. Compte tenu de (7.15), on peut donc affirmer que :

$$E(Z_t|I_t) = d + \Pi Z_t,$$

où d est une matrice colonne donnée et Π est une matrice constante définie par blocs par :

$$\Pi = \begin{bmatrix} & \Pi'_0 & \\ \hline 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \hline 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & & \Phi'_X \\ 0 & & & \end{bmatrix}$$

D'après (P2), on sait que Y_{0t} est une fonction affine de \tilde{Y}_t :

$$\begin{aligned} Y_{0t+1} &= \alpha' \tilde{Y}_{t+1} + g \\ &= \alpha_c Y_{ct+1} + \alpha'_{c(c)} [(Y_{it})'_{i=1,2,\dots,n,m}, X'_t]' + g. \end{aligned}$$

Il est immédiat de vérifier que (7.16) entraîne $\alpha_c \neq 0$. En effet (7.17) implique que :

$$\begin{aligned} E_t Y_{it+1} &= Y_{0t} + f_i \\ &= \alpha' \tilde{Y}_t + f_i. \end{aligned} \tag{A.3}$$

Si α_c était nul, alors :

$$E_t Y_{it+1} = \alpha'_{c(c)} [(Y_{it})'_{i=1,2,\dots,n,m}, X'_t]' + f_i$$

ceci pour $i = 1, 2, \dots, n, m$. On aurait ainsi :

$$E[(Y_{it+1})_{i=1,2,\dots,n,m}|I_t]$$

$$E \left\{ (Y_{it+1})_{i=1,2,\dots,n,m} \mid [(Y_{it})'_{i=1,2,\dots,n,m}, X'_t]'_{t \leq t} \right\}$$

ce qui contredit (7.16).

Il est alors possible d'exprimer Y_{ct} comme une fonction affine de Z_t :

$$\begin{aligned} Y_{ct} &= \frac{1}{\alpha_c} Y_{0t} - \frac{1}{\alpha_c} \alpha'_c [(Y_{it})'_{i=1,2,\dots,n,m}, X'_t]' - \frac{1}{\alpha_c} g \\ &= \beta' Z_t + g^* \end{aligned} \tag{A.4}$$

où $\alpha = [\alpha_c, \alpha'_c]'$, $\beta = [(\alpha_c)^{-1}, (\alpha_c^{-1} \alpha'_c)]'$, et g^* constante réelle donnée. (A.4) implique que le processus d'innovation de Y_{ct+1} (et donc de \tilde{Y}_{t+1}) est une fonction linéaire du processus d'innovation de Z_{t+1} . On aura en outre :

$$\begin{aligned} E_t Y_{ct+1} &= \beta' E_t Z_{t+1} + g^* \\ &= \beta' (d + \Pi Z_t) + g^* \\ &= \beta' \Pi \begin{bmatrix} Y_{0t} \\ (Y_{it})_{i=1,2,\dots,n,m} \\ X_t \end{bmatrix} + g^* + \beta' d^* \\ &= \beta' \Pi \cdot \begin{bmatrix} \alpha' \\ 0 \\ \vdots \\ \vdots \\ \text{Id}_{n+k+1} \\ 0 \end{bmatrix} \cdot \tilde{Y}_t + \beta' \Pi \cdot \begin{bmatrix} g \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} + g^* + \beta' d \\ &= b_c + \Phi'_c \tilde{Y}_t. \end{aligned}$$

Par ailleurs, on sait d'après (A.3) que :

$$E_t Y_{it+1} = \alpha' \tilde{Y}_t + f_i$$

et puisque :

$$E_t X_{t+} = \Phi'_X [(Y_{it})'_{i=1,2,\dots,n,m}, X'_t]' + d_X$$

on peut conclure que :

$$E_t \tilde{Y}_{t+1} = b + \Phi \tilde{Y}_t$$

avec b vecteur colonne réel donné, et Φ définie par blocs par :

$$\Phi = \begin{bmatrix} \Phi'_c \\ \alpha' \\ \vdots \\ \alpha' \\ 0 \\ \vdots \\ 0 \end{bmatrix} \cdot \begin{bmatrix} \Phi'_X \\ \vdots \\ \Phi'_X \end{bmatrix}.$$

Pour obtenir la propriété (P2) il nous reste à vérifier que $\text{Var}_t \tilde{Y}_{t+1}$ est une matrice constante régulière. D'après (A.4) et (P2), toutefois, on sait que :

$$\begin{aligned} \text{Var}_t \tilde{Y}_{t+1} &= \text{Var}_t \begin{bmatrix} \beta' Z_{t+1} \\ (Y_{it+1})_{i=1,2,\dots,n,m} \\ X_{t+1} \end{bmatrix} \\ &= \begin{bmatrix} \beta' \\ 0 \\ \vdots \\ \text{Id}_{n+K+1} \\ 0 \end{bmatrix} \cdot \Omega \cdot \begin{bmatrix} \beta' \\ 0 \\ \vdots \\ \text{Id}_{n+K+1} \\ 0 \end{bmatrix}' . \end{aligned}$$

Cette matrice est régulière, car le premier coefficient de β est non nul (il est en fait donné par $(\alpha_c)^{-1}$). D'où (P1), le caractère VAR(1) gaussien du processus \tilde{Y}_t et la relation annoncée entre les innovations. \square

ANNEXE B

Preuve du Théorème 7.4

Avec des notations évidentes, les espérances et variances de Y_{t+1} conditionnelles à I_t s'écrivent :

$$m_t = \begin{bmatrix} m_{1t} \\ d_2 + B_{21}m_{1t} \end{bmatrix},$$

$$\Sigma_t = \begin{bmatrix} \Sigma_{11t} & \Sigma_{11t}B'_{21} \\ B_{21}\Sigma_{11t} & B_{21}\Sigma_{11t}B'_{21} + \Omega_{22} \end{bmatrix};$$

la performance potentielle de l'ensemble des actifs de base vaut :

$$\begin{aligned} S_t^2 &= m'_t \Sigma_t^{-1} m_t \\ &= [m'_{1t}, d'_2 + m'_{1t}B'_{21}] \cdot \begin{bmatrix} \Sigma_{11t}^{-1} + B'_{21}\Omega_{22}^{-1}B_{21} & -B'_{21}\Omega_{22}^{-1} \\ -\Omega_{22}^{-1}B_{21} & \Omega_{22}^{-1} \end{bmatrix} \cdot \begin{bmatrix} m_{1t} \\ d_2 + B_{21}m_{1t} \end{bmatrix} \\ &= m'_{1t}\Sigma_{11t}^{-1}m_{1t} + d'_2\Omega_{22}^{-1}d_2 \\ &= S_{1t}^2 + d'_2\Omega_{22}^{-1}d_2 \end{aligned}$$

où S_{1t} est la performance potentielle de l'ensemble des K actifs facteurs. Ainsi bien que chacune des mesures de performance S_t et S_{1t} dépende généralement du temps, le modèle est tel que l'écart entre les carrés de ces deux mesures soit stable dans le temps. En particulier les K premiers actifs engendrent l'ensemble des portefeuilles efficients si et seulement si $S_t = S_{1t}, \forall t \Leftrightarrow d_2 = 0$.

Le même type de calcul peut être développé de façon non conditionnelle. Nous notons :

$$m = E m_t,$$

$$\Sigma = E(\Sigma_t) + Var(m_t).$$

Nous voyons que :

$$m = \begin{bmatrix} m_1 \\ d_2 + B_{21}m_1 \end{bmatrix},$$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{11}B'_{21} \\ B_{21}\Sigma_{11} & B_{21}\Sigma_{11}B'_{21} + \Omega_{22} \end{bmatrix};$$

Nous en déduisons que les performances marginales satisfont aussi :

$$S^2 = S_1^2 + d'_2\Omega_{22}^{-1}d_2.$$

□

CHAPITRE 8

LES PROCESSUS ARCH COMME APPROXIMATIONS DE PROCESSUS EN TEMPS CONTINU

par

Laure ELIE

8.1 INTRODUCTION

Les modèles en temps continu se sont largement développés en finance. On peut invoquer plusieurs raisons :

- ces modèles permettent des calculs plus élégants et plus rapides (calcul stochastique, formule d'Itô.)
- ils se prêtent à une théorie simple pour l'évaluation des options comme en témoigne le célèbre modèle de Black and Scholes.
- les cotations des titres sur le marché sont soumises à des fluctuations très fréquentes de telle sorte qu'il est difficile de déterminer l'intervalle de temps adéquat entre deux observations. D'ailleurs les cotations en continu font leur apparition sur les marchés.

Dans ce chapitre il s'agit d'étudier les modèles à temps continu dont les processus discrets de type hétéroscédastique sont de bonnes approximations lorsque l'intervalle de temps entre deux observations tend vers zéro. Les principaux résultats dans ce domaine sont dus à D.B. Nelson [1990a].

Regardons sur des exemples les analogies entre une approche en *temps discret* et une approche en *temps continu*.

Considérons un système déterministe en temps continu dont l'évolution est régie par une équation différentielle ordinaire

$$\begin{cases} \frac{dx_t}{dt} = f(t, x_t) \\ x_0 = x. \end{cases} \quad (8.1)$$

Si le système est observé à des instants successifs de pas h , un modèle discret approché y_n des observations x_{nh} peut être défini par l'équation de récurrence suivante

$$\begin{cases} y_{n+1} = y_n + f[nh, y_n] \times h \\ \quad = y_n + g(n, y_n) \\ y_0 = x \end{cases}$$

les éléments différentiels étant tout simplement remplacés par les accroissements.

Ce modèle discret sera considéré comme une bonne approximation du modèle en temps continu défini en (8.1) s'il « converge » lorsque le pas h tend vers 0 vers ce modèle. Ce sera bien sûr le cas ici si la fonction f est suffisamment régulière.

Supposons maintenant qu'à chaque instant, ce modèle discret soit perturbé par un aléa. Le cas le plus simple consiste à considérer que ce processus perturbateur est un bruit blanc strict, c'est à dire une suite de variables aléatoires ε_n indépendantes et de même loi et que le modèle perturbé s'écrit

$$Y_{n+1} = Y_n + g(n, Y_n) + \varepsilon_{n+1}.$$

Comment représenter un tel processus en temps continu ? Lorsque $g \equiv 0$, le processus $Y_n = Y_0 + \varepsilon_1 + \dots + \varepsilon_n$ est la promenade aléatoire somme des divers aléas. Un processus analogue en temps continu doit refléter en un temps très court l'addition de nombreux chocs indépendants. On peut s'attendre d'après le théorème central limite à ce qu'il soit gaussien. Il serait aussi agréable par similitude avec les promenades aléatoires que ce processus W_t satisfasse la propriété suivante : pour tous $0 \leq s \leq t$, la variable aléatoire $W_t - W_s$, c'est à dire le déplacement sur l'intervalle de temps (s, t) est indépendante de la trajectoire jusqu'à l'instant s , et sa loi ne dépend que de $t - s$, la longueur de l'intervalle de temps considéré. L'exemple typique est le *mouvement brownien* introduit en 1828 par le botaniste Brown pour modéliser la trajectoire continue, mais extrêmement irrégulière avec des changements de direction perpétuels d'un grain de pollen en suspension dans un liquide. Les premières modélisations mathématiques du mouvement brownien remontent à Bachelier [1900] qui l'appliqua aux mouvements des cours boursiers et à Einstein [1905] qui construisit un modèle probabiliste de

diffusion pour décrire ce processus. C'est aussi en 1905 que le physicien Smoluchowski décrit le mouvement brownien comme limite de promenades aléatoires.

On définit alors par analogie au modèle discret le modèle dynamique perturbé en temps continu

$$dX_t = f(t, X_t)dt + \sigma dW_t \tag{8.2}$$

ce qui veut dire pour tout $t \in \mathbb{R}^+$

$$X_t = X_0 + \int_0^t f(s, X_s) ds + \sigma W_t;$$

le processus W_t désigne un mouvement brownien et σ une constante mesurant l'écart type de la perturbation ; c'est le paramètre de *volatilité*.

Supposer la volatilité σ constante, revient à admettre que la perturbation a toujours même variance ce qui correspond dans le cas discret à la modélisation de l'aléa par un bruit blanc strict.

Si l'aléa discret est un *modèle hétéroscédatique*, sa variance ne sera plus constante, et la volatilité σ va devenir aléatoire dans l'équation différentielle stochastique (8.2), comme nous le verrons par la suite.

Nous allons rappeler brièvement les propriétés du mouvement brownien et du calcul stochastique d'Itô.

8.2 CALCUL STOCHASTIQUE

8.2.1 Mouvement Brownien

Définition 8.1

Soit (Ω, B, P) un espace probabilisé. Un processus $(W_t, t \in \mathbb{R}^+)$ est un mouvement brownien si

- il est à accroissements *indépendants* : pour $s \leq t$, $W_t - W_s$ est indépendant de la tribu $\sigma(W_u, u \leq s)$,
- il est à accroissements *stationnaires* : pour $s \leq t$ la loi de $W_t - W_s$ est une loi gaussienne centrée de variance $t - s$,
- $W_0 = 0$ p.s.

Propriétés 8.1

1. Le mouvement brownien est un processus gaussien :

$$\forall t_1, t_2, \dots, t_n \text{ le vecteur } (W_{t_1}, \dots, W_{t_n}) \text{ est gaussien.}$$

2. Pour tous s et t , $\text{cov}(W_t, W_s) = \inf(t, s)$.

3. Le mouvement brownien $(W_t, t \in \mathbb{R}^+)$ admet une version continue : il existe un processus $(W_t^*, t \in \mathbb{R}^+)$ tel que

- $\forall t W_t = W_t^* \text{ p.s.}$
- W_t^* soit un processus à trajectoires continues, c.a.d. tel que, pour tout ω de Ω , l'application $t \rightarrow W_t^*(\omega)$ soit une application continue.

Remarque

Un processus continu à accroissements stationnaires et indépendants est nécessairement un mouvement brownien. Cette propriété est remarquable et est délicate à établir : la stationnarité et l'indépendance des accroissements associées à la continuité du processus caractérisent sa loi.

8.2.2 Intégrale stochastique

Soient (Ω, B, P) un espace probabilisé, $(W_t)_{t \in \mathbb{R}^+}$ un mouvement brownien et $F = (F_t)_{t \in \mathbb{R}^+}$ la filtration naturelle associée : pour tout t , F_t est la tribu engendrée par les variables aléatoires $(W_s, s \leq t)$.

Comme pour l'intégrale de Riemann, l'intégrale stochastique sur $[0, t]$ est d'abord définie pour des processus élémentaires en escalier, c'est à dire de la forme

$$\phi(t, \omega) = \sum_{i=0}^{n-1} X_i(\omega) \mathbf{1}_{]t_i, t_{i+1}]}(t)$$

où $n \in \mathbb{N}$, $0 \leq t_0 < t_1 < t_2 < \dots < t_n \leq T$ et où X_i est une variable aléatoire réelle F_{t_i} -mesurable et de carré intégrable. Un tel processus est adapté à la filtration F au sens où pour tout $t \in [0, T]$, la variable $\phi(t) = \phi(t, \cdot)$ est F_t -mesurable.

L'intégrale stochastique d'Itô de ϕ sur $[0, T]$ est alors par définition

$$I(\phi) = \int_0^T \phi(t) dW_t = \sum_{i=0}^{n-1} X_i (W_{t_{i+1}} - W_{t_i}).$$

L'indépendance de l'accroissement $W_{t_{i+1}} - W_{t_i}$ et de la tribu F_{t_i} permet d'évaluer aisément les moments d'ordre 1 et 2 de $I(\phi)$:

$$\begin{cases} E(I(\phi)) = 0 \\ E(I(\phi)^2) = E \left[\int_0^T \phi(t)^2 dt \right]. \end{cases}$$

L'espace H_T des processus adaptés ϕ définis sur $[0, T] \times \Omega$ tels que

$$E \left[\int_0^T \phi(t)^2 dt \right]$$

soit fini est alors un espace de Hilbert pour la norme définie par

$$\| \phi \| = E \left[\int_0^T \phi(t)^2 dt \right]$$

et l'application $\phi \rightarrow I(\phi)$ est isométrique du sous-espace des processus en escaliers de H_T dans l'espace $L^2(\Omega)$ des variables de carré intégrable.

Nous pouvons donc prolonger l'isométrie au sous-espace fermé de H_T engendré par les processus en escalier. Tout processus φ adapté borné à trajectoires continues est limite de processus en escalier adaptés puisque

$$\varphi = \lim_{n \rightarrow \infty} \sum_{i=0}^{i=n-1} \varphi(iT/n) 1_{]iT/n, (i+1)T/n]}$$

et nous obtenons le résultat suivant :

Proposition 8.1

Soit φ un processus adapté borné à trajectoires continues, l'intégrale stochastique d'Itô de φ sur $[0, T]$ est définie dans $L^2(\Omega)$ par

$$I(\varphi) = \int_0^T \varphi(t) dW_t = \lim_{n \rightarrow \infty} \sum_{i=0}^{i=n-1} \varphi(iT/n) (W_{(i+1)T/n} - W_{iT/n}). \tag{8.3}$$

Elle satisfait

$$\begin{cases} E(I(\varphi)) = 0 \\ E(I(\varphi)^2) = E \left[\int_0^T \varphi(t)^2 dt \right]. \end{cases} \tag{8.4}$$

Remarques :

1. Si le processus à intégrer φ est une fonction déterministe (fonction seulement du temps t), alors la variable $I(\varphi)$ s'écrit comme limite d'après (8.4) d'une suite de variables gaussiennes ; elle est donc elle-même gaussienne centrée de variance

$$\int_0^T \varphi(t)^2 dt.$$

De plus si φ déterministe est continûment différentiable sur \mathbb{R}^+ , nous avons la formule d'intégration par parties suivante qui permet d'exprimer l'intégrale d'Itô $I(\varphi)$ comme intégrale classique par rapport à la mesure de Lebesgue :

$$I(\varphi) = \int_0^T \varphi(t) dW_t = \varphi(T)W_T - \int_0^T \varphi'(t)W_t dt.$$

2. L'intégrale d'Itô a de manière générale la propriété remarquable suivante : pour $t \in [0, T]$,

$$M_t = \int_0^t \varphi(s) dW_s \text{ est une } F_t\text{-martingale de carré intégrable.}$$

Plus précisément :

$$\begin{cases} \mathbb{E} \left(\int_t^{t+h} \varphi(s) dW_s \mid F_t \right) = 0 \\ \text{var} \left(\int_t^{t+h} \varphi(s) dW_s \mid F_t \right) = \mathbb{E} \left(\int_t^{t+h} \varphi(s)^2 ds \right). \end{cases} \quad (8.5)$$

Cette propriété découle de la manière d'approximer l'intégrand par un processus en escalier dont la valeur sur $]t_i, t_{i+1}]$ est sa valeur en t_i borne gauche de l'intervalle. La fonction à intégrer sur $]t_i, t_{i+1}]$ est donc connue sur l'intervalle puisqu'elle est F_{t_i} -mesurable.

Si nous approximations l'intégrand sur l'intervalle $]t_i, t_{i+1}]$ par la moyenne $(\varphi(t_i) + \varphi(t_{i+1}))/2$ des valeurs du processus en t_i et t_{i+1} , nous obtiendrions l'intégrale de Stratanovitch qui n'a pas la propriété de martingale, mais qui obéit aux règles du calcul différentiel, ce qui n'est pas le cas pour l'intégrale d'Itô, comme nous allons le voir.

Enfin l'intégrale d'Itô s'étend aux processus continus non bornés de H_T par troncation.

8.2.3 Processus d'Itô

Nous appellerons processus d'Itô tout processus de la forme

$$X(t) = X(0) + \int_0^t \Psi(s) ds + \int_0^t \varphi(s) dW_s$$

où φ et Ψ sont des processus à trajectoires continues de H_T ($T \geq t$). Sous forme différentielle ce processus s'écrit

$$dX(t) = \Psi(t)dt + \varphi(t)dW_t.$$

Soit $F(t, y)$ une fonction bornée deux fois différentiable de $\mathbb{R}^+ \times \mathbb{R}$ à valeurs réelles, la formule d'Itô va nous permettre d'explicitier le processus $F(t, X(t))$ lorsque $X(t)$ est un processus d'Itô de la forme ci-dessus.

Formule d'Itô :

Le processus $F(t, X(t))$ est un processus d'Itô et s'écrit :

$$\begin{aligned}
 dF(t, X(t)) &= \frac{\partial F}{\partial t}(t, X(t))dt + \frac{\partial F}{\partial y}(t, X(t))dX(t) + \frac{1}{2} \frac{\partial^2 F}{\partial^2 y}(t, X(t))\phi(t)^2 dt \\
 &= \left(\frac{\partial F}{\partial t}(t, X(t)) + \frac{\partial F}{\partial y}(t, X(t))\Psi(t) + \frac{1}{2} \frac{\partial^2 F}{\partial^2 y}(t, X(t))\phi(t)^2 \right) dt \\
 &\quad + \frac{\partial F}{\partial y}(t, X(t))\phi(t)dW_t. \tag{8.6}
 \end{aligned}$$

Il s'ajoute donc un terme supplémentaire par rapport aux règles usuelles de différentiation, le terme

$$\frac{1}{2} \frac{\partial^2 F}{\partial^2 y}(t, X(t))\phi(t)^2 dt$$

spécifique du calcul d'Itô. Une explication heuristique peut être la suivante : l'accroissement

$$dW_t = W_{t+dt} - W_t$$

suivant une loi gaussienne centrée d'écart type $(dt)^{1/2}$ peut être considéré comme un terme d'ordre $(dt)^{1/2}$; si nous faisons un développement limité à l'ordre 2 de $F(t, W_t)$, nous obtenons

$$\begin{aligned}
 dF(t, W_t) &\cong \frac{\partial F}{\partial t}(t, W_t)dt + \frac{\partial F}{\partial y}(t, W_t)dW_t \\
 &\quad + \frac{1}{2} \left(\frac{\partial^2}{\partial t^2}(t, W_t)(dt)^2 + \frac{\partial^2 F}{\partial^2 y}(t, W_t)[dW_t]^2 + 2 \frac{\partial^2 F}{\partial t \partial y}(t, W_t)dt dW_t \right).
 \end{aligned}$$

Si nous gardons seulement les termes d'ordre $(dt)^{1/2}$ et dt , nous avons :

$$dF(t, W_t) \cong \frac{\partial F}{\partial t}(t, W_t)dt + \frac{\partial F}{\partial y}(t, W_t)dW_t + \frac{1}{2} \frac{\partial^2 F}{\partial^2 y}(t, W_t)[dW_t]^2$$

Il suffit de remplacer $[dW_t]^2$ par son espérance dt pour obtenir la formule d'Itô lorsque $X(t) = W_t$. Le cas général s'obtient de la même façon en faisant un développement limité à l'ordre 2 et en respectant les règles :

$$[dW_t]^2 = dt, \quad dt dW_t = 0.$$

8.3 EQUATIONS DIFFERENTIELLES STOCHASTIQUES ET DIFFUSIONS

8.3.1 Définition et propriétés

Définition 8.2

Soient (Ω, B, P) un espace probabilisé, $(W_t)_{t \in \mathbb{R}^+}$ un mouvement brownien et σ, b deux fonctions réelles continues définies sur \mathbb{R} . Un processus $(Y(t))_{t \in [0, T]}$ est solution de l'équation différentielle stochastique :

$$\begin{cases} dY(t) = b(Y(t))dt + \sigma(Y(t))dW_t & (t \leq T) \\ Y(0) = y_0 \end{cases} \quad (8.7)$$

si et seulement si $Y(t)$ est un processus adapté continu satisfaisant pour $t \leq T$

$$Y(t) = y_0 + \int_0^t b(Y(s))ds + \int_0^t \sigma(Y(s))dW_s.$$

Il y a *existence locale* et unicité en loi de la solution dans les deux cas suivants :

– soit les fonctions σ et b sont *localement lipschitziennes*, (8.8)

– soit la fonction $\sigma(y)$ est *strictement positive pour tout y* . (8.9)

Pour obtenir l'*existence globale* sur $[0, T]$, il faut imposer à σ et b d'être sous-linéaires, c.a.d.

$$\exists K > 0, \forall y \in \mathbb{R} \quad |\sigma(y)| + |b(y)| \leq K(1 + |y|). \quad (8.10)$$

Propriétés 8.2

1. La solution $Y(t)$ appelée *processus de diffusion* est un processus markovien.
2. Le terme $\sigma(y)$ s'appelle *coefficient de diffusion* ou *volatilité* et le terme $b(y)$ *coefficient de dérive* ou *tendance*. Ils vérifient :

$$\lim_{h \rightarrow 0^+} h^{-1} \mathbb{E}(Y(t+h) - Y(t) | Y(t) = y) = b(y) \quad (8.11)$$

$$\lim_{h \rightarrow 0^+} h^{-1} \text{var}(Y(t+h) - Y(t) | Y(t) = y) = \sigma^2(y). \quad (8.12)$$

En effet

$$Y(t+h) - Y(t) = \int_t^{t+h} b(Y(s))ds + \int_t^{t+h} \sigma(Y(s))dW_s,$$

d'où, d'après (8.5)

$$E(Y(t+h) - Y(t)) | Y(t) = y = E \left(\int_t^{t+h} b(Y(s)) ds | Y(t) = y \right)$$

et

$$\text{var}(Y(t+h) - Y(t)) | Y(t) = y = E \left(\int_t^{t+h} \sigma(Y(s))^2 ds | Y(t) = y \right).$$

Il suffit d'utiliser la continuité en s de $\sigma(Y(s))$ et de $b(Y(s))$ pour conclure.

3. La diffusion $Y(t)$ a la propriété de martingale suivante :

Si φ est une fonction deux fois continûment différentiable à support compact, alors le processus

$$\varphi(Y(t)) - \varphi(Y(0)) - \int_0^t L\varphi(Y(s)) ds,$$

où L est l'opérateur différentiel

$$L = \frac{1}{2} \sigma^2 \frac{\partial^2}{\partial y^2} + b \frac{\partial}{\partial y} \tag{8.13}$$

est une F_t -martingale.

On applique la formule d'Itô à $\varphi''(Y(t))$. Il vient :

$$\begin{aligned} d\varphi(Y(t)) &= \varphi'(Y(t)) dY(t) + \frac{1}{2} \varphi''(Y(t)) \sigma^2(Y(t)) dt \\ &= \left(b(Y(t)) \varphi'(Y(t)) + \frac{1}{2} \sigma^2(Y(t)) \varphi''(Y(t)) \right) dt \\ &\quad + \varphi'(Y(t)) \sigma(Y(t)) dW_t. \end{aligned}$$

Il suffit de remarquer que l'intégrale stochastique

$$\int_0^t \varphi'(Y(s)) \sigma(Y(s)) dW_s$$

est une martingale pour conclure.

Cette formulation en termes de martingales est intéressante car on peut montrer que réciproquement si $Y(t)$ satisfait (8.13) pour une classe de fonctions φ suffisamment régulières, alors $Y(t)$ est solution de l'EDS ci-dessus. On a donc une caractérisation de la solution $Y(t)$ de l'EDS à l'aide du générateur L .

Une conséquence de l'assertion (8.13) ci-dessus est que

$$E(\varphi(Y(t))) = E(\varphi(Y(0))) + \int_0^t E(L\varphi(Y(s))) ds. \tag{8.14}$$

8.3.2 Exemples de diffusions

Comme dans le cas déterministe, il est en général difficile de décrire explicitement les solutions des équations différentielles stochastiques (EDS). Nous allons passer en revue quelques cas simples, les EDS linéaires, où le calcul explicite est possible.

Exemple 1 : Le brownien géométrique.

Soient b et σ des constantes > 0 ; on considère l'EDS

$$\begin{cases} dS(t) = S(t)(bdt + \sigma dW_t) \\ S(0) = s_0 \end{cases} \quad (s_0 > 0) \quad (8.15)$$

Nous savons d'après (8.8) qu'il existe une solution $S(t)$. Posons $Y(t) = \ln(S(t))$ sur le domaine où $S(t)$ reste positif. D'après la formule d'Itô,

$$dY(t) = \left(b - \frac{1}{2}\sigma^2\right) dt + \sigma dW_t$$

et donc

$$\ln(S(t)) = \ln s_0 + \left(b - \frac{1}{2}\sigma^2\right)t + \sigma W_t,$$

c.a.d.

$$S(t) = s_0 \exp\left(\left(b - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right). \quad (8.16)$$

La solution est donc toujours strictement positive et le passage au logarithme était donc justifié. La loi de $S(t)$ est lognormale et $\ln(S(t))$ a pour variance $\sigma^2 t$. Ce processus a la propriété suivante : si $0 \leq t_1 \leq t_2 \leq \dots \leq t_n$ avec pour $i \in \{2, \dots, n\}$, $t_i - t_{i-1} = \Delta$, les v.a. $\ln(S(t_i)/S(t_{i-1}))$ sont, pour $i \in \{2, \dots, n\}$, indépendantes et de même loi gaussienne d'espérance

$$\left(b - \frac{1}{2}\sigma^2\right)\Delta$$

et de variance $\sigma^2 \Delta$, par suite les accroissements multiplicatifs $S(t_i)/S(t_{i-1})$ sont indépendants et de même loi. C'est cette dernière assertion qui conduit à modéliser *le cours des actifs* par un *brownien géométrique*, les *rendements* $(S(t_i) - S(t_{i-1}))/S(t_{i-1})$ de l'actif étant considérés indépendants et de même loi. Ce modèle proposé par Black & Sholes est utilisé couramment pour le cours des actifs. Il présuppose que les perturbations du rendement des actifs ont à des instants équadistribués une variance constante.

Exemple 2 : Le processus d'Ornstein-Uhlenbeck.

Soient a, b, σ des constantes ; on considère l'EDS

$$\begin{cases} dY(t) = (\omega - \theta Y(t))dt + \sigma dW_t \\ Y(0) = y_0 \end{cases} \quad (8.17)$$

Posons $Z(t) = Y(t) \exp(\theta t)$; alors d'après la formule d'Itô :

$$\begin{aligned} dZ(t) &= \exp(\theta t) [(\omega - \theta Y(t))dt + \sigma dW_t] + Y(t)(\theta \exp(\theta t)) \\ &= \omega \exp(\theta t)dt + \exp(\theta t)\sigma dW_t. \end{aligned}$$

Par suite

$$Z(t) = Z(0) \left(\int_0^t \omega \exp(\theta s)ds + \int_0^t \exp(\theta s)dW_s \right)$$

et donc

$$Y(t) = \frac{\omega}{\theta}(1 - \exp(-\theta t)) + y_0 \exp(-\theta t) + \sigma \int_0^t \exp[\theta(s - t)]dW_s \quad (8.18)$$

Le processus $Y(t)$ s'exprimant à l'aide d'une intégrale stochastique de fonction déterministe est un processus gaussien satisfaisant

$$\begin{aligned} E(Y(t)) &= \frac{\omega}{\theta}(1 - \exp(-\theta t)) + y_0 \exp(-\theta t) \\ \text{cov}(Y(u), Y(t)) &= \int_0^u \exp[\theta(s - u)]\sigma^2 \exp[\theta(s - t)]ds \quad (u \leq t) \\ &= \frac{\sigma^2}{2\theta}(\exp -\theta(t - u) - \exp -\theta(t + u)). \end{aligned}$$

Lorsque t tend vers l'infini, la loi de $Y(t)$ tend, si θ est > 0 , vers une loi normale d'espérance ω/θ et de variance $\sigma^2/2\theta$. Cette loi est une *distribution stationnaire*, c'est une probabilité invariante pour le processus de Markov $Y(t)$.

Ce processus est utilisé pour des séries qui ont un comportement relativement stationnaire oscillant autour d'une moyenne. En finance, les *taux d'intérêt* sont souvent modélisés par un tel processus, le paramètre ω/θ étant la valeur moyenne du taux. Le reproche principal à ce type de modèle pour un taux est qu'étant un modèle gaussien, il peut prendre des valeurs *négatives*. Pour remédier à ce défaut, divers modèles de la forme

$$dY(t) = (\omega - \theta Y(t))dt + \sigma [Y(t)]^\alpha dW_t \quad \alpha \in \left[\frac{1}{2}, 1 \right]$$

ont été introduits pour lesquels un bon choix des paramètres ω, θ, σ assure que la solution $Y(t)$ reste strictement positive. Nous allons expliciter le cas $\alpha = 1$.

Exemple 3 : Le modèle linéaire.

Soient ω, θ, σ des constantes ; on considère l'EDS

$$\begin{cases} dY(t) = (\omega - \theta Y(t))dt + \sigma Y(t)dW_t \\ Y(0) = y_0 \end{cases} \quad (8.19)$$

Le cas $\omega = 0$ correspond au brownien géométrique et nous avons vu qu'alors la solution était de la forme

$$\xi(t) = \exp\left(-\left(\theta + \frac{1}{2}\sigma^2\right)t + \sigma W_t\right).$$

Nous utilisons la méthode de la « variation de la constante » pour trouver $Y(t)$. Posons $Z(t) = Y(t)\xi(t)^{-1}$; alors d'après la formule d'Itô,

$$d(\xi(t)^{-1}) = \xi(t)^{-1}((\theta + \sigma^2)dt - \sigma dW_t)$$

et

$$\begin{aligned} dZ(t) &= Y(t)d(\xi(t)^{-1}) + \xi(t)^{-1}dY(t) - \sigma^2 Y(t)\xi(t)^{-1}dt \\ &= \xi(t)^{-1}\omega dt. \end{aligned}$$

Donc

$$Y(t) = \xi(t) \left[y_0 + \omega \int_0^t \xi(s)^{-1} ds \right]. \quad (8.20)$$

On remarque alors que, si ω et y_0 sont positifs ou nuls avec $(\omega, y_0) \neq 0$, la solution reste *strictement positive*.

Dans la mesure où nous allons rechercher les processus en temps continu dont les modèles hétéroscédastiques sont des approximations, il est intéressant de déterminer des diffusions qui admettent une loi stationnaire puisque les modèles discrets considérés seront souvent *stationnaires*.

8.3.3 Probabilité invariante

Soit l'EDS

$$dY(t) = b(Y(t))dt + \sigma(Y(t))dW_t,$$

où b et σ sont des fonctions continues et où $\sigma(y)$ est strictement positive pour $y \in]\alpha, \beta[$. Supposons que le processus $Y(t)$ reste dans l'intervalle $]\alpha, \beta[$. Soit $c \in]\alpha, \beta[$, considérons la fonction g définie sur $]\alpha, \beta[$ par

$$g(y) = \frac{1}{\sigma^2(y)} \exp\left(2 \int_c^y \frac{b(x)}{\sigma^2(x)} dx\right). \quad (8.21)$$

Alors le processus de Markov admet une probabilité invariante si et seulement si g est intégrable sur (α, β) et cette probabilité invariante unique s'écrit

$$m(dy) = \frac{g(y)}{\gamma} dy \tag{8.22}$$

où $\gamma = \int_{\alpha}^{\beta} g(y)dy$.

Vérifions qu'une telle probabilité est invariante. D'après (8.14) si φ est une fonction C^2 à support compact,

$$E(\varphi(Y(t))) = E(\varphi(Y(0))) + \int_0^t E(L\varphi(Y(s))) ds.$$

Par suite si m est une loi stationnaire pour la diffusion $Y(t)$, elle est portée par (α, β) et satisfait

$$\int_0^t \int_{\alpha}^{\beta} L\varphi(y)m(dy)ds = 0 \quad \forall t \in \mathbb{R}$$

et donc

$$\int_{\alpha}^{\beta} L\varphi(y)m(dy) = 0.$$

Par suite si m a une densité g , m sera une loi stationnaire si et seulement si

$$L^*g = 0$$

où L^* est l'opérateur adjoint de L , c.a.d. vaut

$$\frac{1}{2}\sigma^2 \frac{\partial^2}{\partial y^2} - b \frac{\partial}{\partial y}.$$

Il suffit alors de vérifier que la fonction g définie en (8.21) vérifie $L^*g = 0$ sur (α, β) .

Application

Considérons le modèle linéaire (exemple 3) ; la solution $Y(t)$ reste dans $]0, \infty[$ si ω et y_0 sont ≥ 0 et tels que $(\omega, y_0) \neq (0, 0)$; de plus pour $y \in]0, \infty[$, $\sigma(y) = \sigma y$ est strictement positif. Prenons $c = 1$; alors

$$\begin{aligned} g(y) &= \frac{1}{\sigma^2 y^2} \exp\left(2 \int_1^y \frac{\omega - \theta x}{\sigma^2 x^2} dx\right) \\ &= \frac{\exp\left(\frac{2\omega}{\sigma^2}\right)}{\sigma^2} \exp\left(\frac{-2\omega}{\sigma^2 y}\right) y^{-2[(\theta/\sigma^2)+1]}. \end{aligned}$$

Il est facile de voir que, pour $\omega > 0$, g est intégrable sur \mathbb{R}^+ si et seulement si

$$-\frac{2\theta}{\sigma^2} < 1. \quad (8.23)$$

La probabilité invariante a dans ce cas pour densité sur \mathbb{R}^+

$$\left(\frac{2\omega}{\sigma^2}\right)^{[(2\theta)/\sigma^2]+1} \Gamma\left(\frac{2\theta}{\sigma^2} + 1\right)^{-1} \exp\left(-\frac{2\omega}{\sigma^2}y\right) y^{-2[(\theta/\sigma^2)+1]} \quad (8.24)$$

Γ désignant la fonction gamma. Si $Y(0)$ a pour loi cette probabilité, alors le processus $Y(t)$ est stationnaire et $Y(t)^{-1}$ suit une *loi gamma*

$$\Gamma\left(\frac{2\omega}{\sigma^2}, \frac{2\theta}{\sigma^2} + 1\right).$$

On rappelle que la loi $\Gamma(\lambda, a)$ est définie pour λ et a positifs et a pour densité sur \mathbb{R}^+ :

$$\frac{\lambda^a}{\Gamma(a)} \exp(-\lambda y) y^{a-1}.$$

8.3.4 Le cas vectoriel

On appelle mouvement brownien d -dimensionnel un processus indexé par \mathbb{R}^+ , $W_t = (W_t^1, W_t^2, \dots, W_t^d)$ tel que les processus W_t^i , $i \in \{1, 2, \dots, d\}$, soient des mouvements browniens réels indépendants. Soit $F = (F_t)_{t \in \mathbb{R}^+}$ où F_t est la tribu engendrée par les v.a. W_s pour $s \leq t$, la filtration naturelle associée.

Si ϕ un processus défini sur $[0, T]$, continu adapté à la filtration F , et à valeurs matrices $p \times d$, on définit l'intégrale d'Itô de ϕ

$$\int_0^T \phi(t) dW_t$$

comme le vecteur aléatoire de dimension p dont la i -ème coordonnée est donnée par

$$\sum_{j=1}^d \int_0^T \phi_{ij}(t) dW_t^j.$$

Nous pouvons alors considérer, si σ (respectivement b) est une application continue définie sur \mathbb{R}^p à valeurs matrices $p \times d$ (resp. à valeurs \mathbb{R}^p), l'équation différentielle stochastique

$$\begin{cases} dY(t) = b(Y(t))dt + \sigma(Y(t))dW_t & (t \leq T) \\ Y(0) = y_0 \end{cases}$$

où la solution sera un processus à valeurs \mathbb{R}^p . Les conditions d'existence et d'unicité sont analogues au cas scalaire. La propriété (8.12) sur la variance de l'accroissement se réécrit

$$\lim_{h \rightarrow 0^+} h^{-1} \text{var}[Y(t+h) - Y(t)|Y(t) = y] = \sigma(t, y)\sigma(t, y)' \quad (8.25)$$

où $\text{var } Z$ désigne la matrice de covariance $E[(Z - E(Z))(Z - E(Z))']$.

8.4 CONVERGENCE DE CHAINES DE MARKOV VERS DES DIFFUSIONS

8.4.1 Discrétisation d'Euler d'une diffusion

Soit sur \mathbb{R}^p l'EDS

$$\begin{cases} dY(t) = b(Y(t))dt + \sigma(Y(t))dW_t & (t \leq T) \\ Y(0) = y_0 \end{cases}$$

où W_t est un brownien d -dimensionnel.

Comme dans le cas déterministe, une discrétisation naturelle de la solution s'obtient en remplaçant les *éléments différentiels par leurs accroissements*. La *discrétisation d'Euler* de pas h de la diffusion $Y(t)$ solution de l'EDS est justement le processus $(Y_{kh})_{k \in \mathbb{N}}$ satisfaisant

$$\begin{cases} Y_{(k+1)h} = Y_{kh} + h b(Y_{kh}) + \sqrt{h} \sigma(Y_{kh})z_{k+1}^{(h)} \\ Y_0 = y_0 \end{cases}$$

où les v.a. $z_k^{(h)}$ ($K \in \mathbb{N}$) sont des vecteurs gaussiens de \mathbb{R}^d indépendants, centrés et de matrice de covariance la matrice identité de \mathbb{R}^d .

Ce processus Y_{kh} est une chaîne de Markov qui vérifie

$$\begin{cases} h^{-1} E(Y_{(k+1)h} - Y_{kh})|Y_{kh} = y = b(y) \\ h^{-1} \text{var}[(Y_{(k+1)h} - Y_{kh})|Y_{kh} = y] = \sigma(y)\sigma(y)' \end{cases} \quad (8.26)$$

Nous allons voir que ce type de propriété analogue à (8.11) et (8.12) va assurer que la suite de chaînes de Markov Y_{kh} indexée par h converge dans un sens que l'on va préciser vers la diffusion $Y(t)$ lorsque $h \rightarrow 0$.

8.4.2 Théorème de convergence

Considérons une famille de chaînes de Markov $(Y_0^{(h)}, Y_h^{(h)}, \dots, Y_{kh}^{(h)}, \dots)$ (indexée par h) à valeurs dans \mathbb{R}^d . La notation $Y_{kh}^{(h)}$, indique que le temps

h sépare deux observations successives. Une telle chaîne de Markov peut se prolonger en un *processus en temps continu* en supposant qu'elle reste constante entre deux observations successives :

$$Y_t^{(h)} = Y_{kh}^{(h)} \quad \text{si} \quad kh \leq t < (k+1)h \quad (k \in \mathbb{N}).$$

On veut étudier la convergence des processus $Y_t^{(h)}$ (qui sont continus à droite et limités à gauche) vers la diffusion $Y(t)$. L'approche la plus simple consiste à étudier la convergence des *lois fini-dimensionnelles* des processus, mais il est aussi possible d'étudier la convergence des lois des processus $Y_t^{(h)}$ considérés globalement sur $[0, T]$ en se plaçant sur l'espace des fonctions CAD-LAG (fonctions continues à droite et limitées à gauche) muni d'une topologie adéquate *la topologie de Skorohod* (Billingsley [1968]). Nous n'insisterons pas sur ce type de convergence, mais le théorème énoncé ci-dessous est vrai dans ce cadre. Les résultats de convergence que nous donnons ici émanent des livres de Stroock-Varadhan [1979] et de Ethier-Kurtz [1986].

Théorème 8.1

Soit une famille de chaînes de Markov $(Y_0^{(h)}, Y_h^{(h)}, \dots, Y_{kh}^{(h)}, \dots)$ (indexée par h) à valeurs dans \mathbb{R}^d . Supposons qu'il existe des applications continues a et b telles que, pour y de \mathbb{R}^d , $a(y)$ soit une matrice d -dimensionnelle définie non négative et $b(y)$ un vecteur de \mathbb{R}^d et telles que, pour tout r de \mathbb{R}^+ , $a(y)$, on ait

- (1) $\sup_{|y| \leq r} \left| h^{-1} \mathbb{E} \left[(Y_h^{(h)} - Y_0^{(h)}) | Y_0^{(h)} = y \right] - b(y) \right| \rightarrow 0$ quand $h \rightarrow 0$.
- (2) $\sup_{|y| \leq r} \left| h^{-1} \text{var} \left[(Y_h^{(h)} - Y_0^{(h)}) | Y_0^{(h)} = y \right] - a(y) \right| \rightarrow 0$ quand $h \rightarrow 0$.
- (3) $\sup_{|y| \leq r} \left| h^{-[1+(\delta/2)]} \mathbb{E} \left[|Y_h^{(h)} - Y_0^{(h)}|^{2+\delta} | Y_0^{(h)} = y \right] \right|$ reste, pour un $\delta > 0$, borné quand $h \rightarrow 0$.

Supposons de plus qu'il existe une application σ continue de \mathbb{R}^d dans l'ensemble des matrices d -dimensionnelles telle que

$$\forall y \in \mathbb{R}^d \quad a(y) = \sigma(y)' \sigma(y)'.$$

Considérons alors l'EDS

$$\begin{cases} dY(t) = b(Y(t))dt + \sigma(Y(t))dW_t \\ Y(0) = y_0. \end{cases}$$

Si cette EDS admet une solution $Y(t)$ unique en loi (par exemple satisfait à (8.8) ou (8.9)) et si $Y_0^{(h)}$ converge en loi vers y_0 , alors les lois fini-dimensionnelles des processus $Y_t^{(h)}$ ($= Y_{kh}^{(h)}$ si $kh \leq t < (k+1)h$) convergent vers celles de la solution $Y(t)$ de l'EDS.

Remarque

Les conditions (1) et (2) sont naturelles et généralisent en quelque sorte (8.11) et (8.12).

La condition (3) assure entre autres que, pour tout $\varepsilon > 0$,

$$h^{-1} P \left(\left| Y_{(k+1)h}^{(h)} - Y_{kh}^{(h)} \right| > \varepsilon \mid Y_{kh}^{(h)} = y \right)$$

tend vers 0 lorsque $h \rightarrow 0$ et par suite que le processus limite n'a pas de sauts.

Application immédiate

La *discrétisation d'Euler* d'une diffusion converge vers cette diffusion sous les conditions (8.8) ou (8.9). En effet d'après (8.26) les conditions (1) et (2) du théorème sont vérifiées; la condition (3) se vérifie facilement sur le moment d'ordre 3 ($\delta = 1$).

En particulier si nous considérons la discrétisation de l'EDS $dY(t) = dW_t$, nous obtenons les *marches aléatoires* $Y_{kh}^{(h)}$ définies par

$$Y_{(k+1)h}^{(h)} - Y_{kh}^{(h)} = \sqrt{h} z_{k+1}^{(h)}$$

où les v.a. $(z_k^{(h)})_{k \in \mathbb{N}}$ sont indépendantes, gaussiennes centrées réduites. Nous retrouvons alors le résultat de *Donsker*, à savoir que la suite de processus

$$Y_t^{(h)} = \sqrt{h} \sum_{i=1}^{[t/h]} z_i^{(h)}$$

converge, lorsque h tend vers 0, vers le mouvement brownien W_t .

Remarque

Le fait que les v.a. $z_k^{(h)}$ soient des vecteurs gaussiens indépendants n'est pas important pour la convergence de la discrétisation d'Euler d'une diffusion. Ce qu'il faut, c'est que les vecteurs aléatoires $z_k^{(h)}$ soient centrés, de matrice de covariance égale à la matrice identité et non corrélés entre eux.

**8.5 LES PROCESSUS HÉTÉROSCEDASTIQUES
COMME APPROXIMATIONS DE PROCESSUS
EN TEMPS CONTINU**

En finance le modèle de diffusion le plus simple pour modéliser le cours $S(t)$ d'un actif est le brownien géométrique (exemple 1 de 8.3.2):

$$dS(t) = S(t)(b dt + \sigma dW_t)$$

où b et σ sont des constantes réelles.

La solution $S(t)$ est strictement positive si la valeur initiale l'est et, posant $Y(t) = \ln S(t)$, nous avons

$$dY(t) = \left(b - \frac{\sigma^2}{2} \right) dt + \sigma dW_t.$$

La discrétisation d'Euler de pas h de ce modèle s'écrit

$$Y_{kh} - Y_{(k-1)h} = \left(b - \frac{\sigma^2}{2} \right) h + \sigma \sqrt{h} z_k \quad (8.27)$$

où les $(z_k)_{k \in \mathbb{N}}$ sont gaussiennes indépendantes centrées réduites.

Ce modèle discret suppose la volatilité σ constante. Nous savons que cette hypothèse est peu conforme à la réalité et nous allons introduire dans le modèle (8.27) une variance conditionnelle non constante de *type GARCH, ARCH exponentiel, ARCH à seuil* et étudier quelle est dans ces divers cas la diffusion limite, s'il en existe une.

Le modèle que nous allons considérer sera donc le suivant :

$$Y_t = Y_{t-1} + f(\sigma_t^2) + \sigma_t z_t \quad (8.28)$$

où les v.a. $(z_t)_{t \in \mathbb{N}}$ sont i.i.d. centrées de variance 1 (non nécessairement gaussiennes) et f est une fonction continue à valeurs réelles. Pour tout t de \mathbb{N} , la v.a. σ_t dépendra du passé du processus z jusqu'à l'instant $t - 1$.

La variable σ_t^2 représente donc la *variance conditionnelle* de la perturbation $\varepsilon_t = \sigma_t z_t$. Divers modèles vont être considérés pour cette variance conditionnelle.

8.5.1 Modèle GARCH-M (D.B. Nelson [1990a])

Supposons pour commencer que la perturbation ε_t soit un processus GARCH; dans ce cas le processus Y_t est un modèle GARCH-M (défini en 3.1.4).

La *variance conditionnelle* σ_t^2 s'écrit donc, si le processus ε_t est GARCH(p, q)

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2,$$

avec $\omega > 0$, $\alpha_i \geq 0$ pour $i \in \{1, \dots, q\}$, $\beta_j \geq 0$ pour $j \in \{1, \dots, p\}$, $\alpha_q \neq 0$, $\beta_p \neq 0$.

Nous nous intéressons à des modèles *markoviens* puisque nous cherchons des modèles limites du *type diffusion*. Nous remarquons que le processus (Y_t, σ_{t+1}^2) est markovien si et seulement si le processus ε_t est GARCH(1,1). Le

modèle s'écrit alors

$$\begin{cases} Y_t &= Y_{t-1} + f(\sigma_t^2) + \sigma_t z_t \\ \sigma_{t+1}^2 &= \omega + \sigma_t^2 (\beta + \alpha z_t^2). \end{cases} \quad (8.29)$$

Considérons une suite de modèles de ce type indexés par h tels que les coefficients ω, β et α dépendent de manière adéquate de h . Au vu de la discrétisation (8.27), nous choisissons la suite de modèles suivants

$$\begin{cases} Y_{(k+1)h} &= Y_{kh} + f(\sigma_{kh}^2) h + \sigma_{kh} \sqrt{h} z_{k+1}^{(h)} \\ \sigma_{(k+1)h}^2 &= \omega_h + \sigma_{kh}^2 (\beta_h + \alpha_h (z_{k+1}^{(h)})^2) \end{cases} \quad (8.30)$$

où les v.a. $z_k^{(h)}$ sont, pour $k \in \mathbb{N}$, i.i.d. centrées de variance 1.

Il s'agit de trouver des suites ω_h, β_h et α_h telles que les conditions (1) et (2) du théorème de convergence 8.1 soient vérifiées. Comme le modèle ci-dessus est bivarié, il en sera de même de la diffusion limite si elle existe.

On remarque que

$$\begin{cases} Y_h - Y_0 &= f(\sigma_0^2) h + \sigma_0 \sqrt{h} z_1^{(h)} \\ \sigma_h^2 - \sigma_0^2 &= \omega_h + \sigma_0^2 (\beta_h + \alpha_h (z_1^{(h)})^2 - 1). \end{cases} \quad (8.31)$$

Il en résulte que pour l'espérance conditionnelle, nous avons

$$h^{-1} \mathbb{E} [(Y_h - Y_0) | (Y_0 = y, \sigma_0^2 = x)] = f(x)$$

et

$$\begin{aligned} &h^{-1} \mathbb{E} [(\sigma_h^2 - \sigma_0^2) | (Y_0 = y, \sigma_0^2 = x)] \\ &= h^{-1} \mathbb{E} \left[\left(\omega_h + \sigma_0^2 (\beta_h + \alpha_h (z_1^{(h)})^2 - 1) \right) \middle| (\sigma_0^2 = x) \right] \\ &= h^{-1} \omega_h + x h^{-1} (\beta_h + \alpha_h - 1). \end{aligned}$$

Cette dernière suite converge pour tout x si et seulement si on a

- (i) $h^{-1} \omega_h \rightarrow \omega$ quand $h \rightarrow 0$
- (ii) $h^{-1} (\beta_h + \alpha_h - 1) \rightarrow -\theta$ quand $h \rightarrow 0$.

Posons $(X_{kh} = (Y_{kh}, \sigma_{kh}^2)'$. Si les conditions (i) et (ii) sont satisfaites, nous obtenons

$$h^{-1} E[(X_h - X_0) | X_0 = (y, x)] \rightarrow b(y, x) = \begin{bmatrix} f(x) \\ \omega - \theta x \end{bmatrix}$$

lorsque h tend vers 0.

Remarque

Il a été vu au chapitre 3 qu'il existe une *solution stationnaire stricte* pour le modèle GARCH(1,1) $\varepsilon_{kh} = \sigma_{(k-1)h} z_k^{(h)}$ si et seulement si

$$E \left[\ln \left(\beta_h + \alpha_h (z_1^{(h)})^2 \right) \right] < 0,$$

et que de plus cette solution a un moment d'ordre 2, et donc σ_{kh}^2 un moment d'ordre 1, si et seulement si

$$\beta_h + \alpha_h < 1.$$

Le moment m_h d'ordre 1 de σ_h^2 satisfait alors

$$m_h = E(\sigma_h^2) = \omega_h + m_h (\beta_h + \alpha_h)$$

et donc

$$m_h = \frac{\omega_h}{1 - \beta_h - \alpha_h}. \quad (8.32)$$

La condition (ii) ci-dessus assure que $\beta_h + \alpha_h$ tend vers 1 lorsque $h \rightarrow 0$; le modèle se rapproche donc de la *frontière entre les zones 1 et 2* de la figure 3.5 du chapitre 3. Si le modèle reste à l'intérieur de la zone 1, c'est à dire si $\beta_h + \alpha_h$ reste toujours inférieur à 1, alors le moment d'ordre 1 de σ_h^2 converge, si (i) et (ii) sont satisfaites, vers $\frac{\omega}{\theta}$.

Considérons maintenant la variance conditionnelle :

$$h^{-1} \text{var}[(Y_h - Y_0) | (Y_0 = y, \sigma_0^2 = x)] = x,$$

$$h^{-1} \text{var}[(\sigma_h^2 - \sigma_0^2) | (Y_0 = y, \sigma_0^2 = x)]$$

$$= h^{-1} \text{var} \left[\left(\sigma_0^2 \alpha_h (z_1^{(h)})^2 \right) | (\sigma_0^2 = x) \right]$$

$$= x^2 \frac{\alpha_h^2}{h} \text{var} \left((z_1^{(h)})^2 \right).$$

Cette suite converge si la condition suivante est vérifiée :

$$(iii) \quad \frac{\alpha_h^2}{h} \rightarrow \alpha \text{ lorsque } h \rightarrow 0.$$

Il résulte alors des conditions (ii) et (iii) que le couple (α_h, β_h) doit tendre vers $(0,1)$ lorsque h tend vers 0 et donc que le modèle se rapproche non seulement de la frontière entre les zones 1 et 2 de la figure 3.5 du chapitre 3 mais aussi du sommet en haut à gauche du rectangle représenté sur cette figure.

Pour terminer, il faut considérer le terme croisé

$$\begin{aligned} & h^{-1} \text{cov}[(Y_h - Y_0), (\sigma_h^2 - \sigma_0^2) | (Y_0 = y, \sigma_0^2 = x)] \\ &= h^{-1} \text{E} \left[\sqrt{x} \sqrt{h} z_1^{(h)} x \alpha_h \left((z_1^{(h)})^2 - 1 \right) \right] \\ &= \frac{\alpha_h}{\sqrt{h}} x^{3/2} \text{E} \left((z_1^{(h)})^3 \right). \end{aligned}$$

Celui-ci converge aussi sous (iii). Par suite si nous posons

$$\begin{aligned} \text{E} \left((z_1^{(h)})^3 \right) &= D \\ \text{E} \left((z_1^{(h)})^4 \right) &= C, \end{aligned}$$

nous obtenons, si (iii) est vérifiée

$$h^{-1} \text{var}[(X_h - X_0) | X_0 = (y, x)] \rightarrow a(y, x) = \begin{bmatrix} x & x^{3/2} \sqrt{\alpha} D \\ x^{3/2} \sqrt{\alpha} D & x^2 \alpha (C - 1) \end{bmatrix}$$

lorsque $h \rightarrow 0$.

La matrice $a(y, x)$ s'écrit aussi $\sigma(y, x)\sigma(y, x)'$ avec

$$\sigma(y, x) = \begin{bmatrix} \sqrt{x} & 0 \\ x\sqrt{\alpha} D & x\sqrt{\alpha} \sqrt{C - 1 - D^2} \end{bmatrix}.$$

Proposition 8.2

Sous les conditions (i, ii, iii), la famille de chaînes de Markov définies en (8.30) et valant (y_0, x_0) avec $x_0 \geq 0$ à l'instant 0 converge (selon le sens défini dans le théorème 8.1), lorsque h tend vers 0, vers la diffusion bivariable

$$\begin{cases} dY_t &= f(\sigma_t^2)dt + \sigma_t dW_t^1 \\ d\sigma_t^2 &= (\omega - \theta\sigma_t^2) dt + \sigma_t^2 \sqrt{\alpha} D dW_t^1 \\ &\quad + \sigma_t^2 \sqrt{\alpha} \sqrt{C - 1 - D^2} dW_t^2 \\ (Y_0, \sigma_0^2) &= (y_0, x_0) \end{cases} \tag{8.33}$$

où W^1 et W^2 désignent deux browniens indépendants.

Les coefficients D et C représentent respectivement les moments d'ordre 3 et d'ordre 4 de la v.a. $z_1^{(h)}$.

Preuve

Il suffit d'appliquer le théorème 8.1. L'uniformité des convergences dans 1) et 2) sur les boules est immédiate. Pour la condition 3), il suffit de remarquer, en utilisant (8.31) et les conditions (i), (ii), (iii), que pour h suffisamment petit, il existe des constantes positives $\lambda, \mu, \lambda', \mu'$ telles que

$$h^{-(1+\delta/2)} \sup_{|x| \leq R} E[(Y_h - Y_0)^{2+\delta} | \sigma_0^2 = x] \leq E[(\lambda + \mu z_1^{(h)})^{2+\delta}]$$

$$h^{-(1+\delta/2)} \sup_{|x| \leq R} E[(\sigma_h^2 - \sigma_0^2)^{2+\delta} | \sigma_0^2 = x] \leq E[(\lambda' + \mu' (z_1^{(h)})^2)^{2+\delta}].$$

De plus l'EDS limite admet une solution unique : en effet σ_t^2 satisfait une EDS linéaire (exemple 3 du paragraphe 8.3.2.) et s'écrit de manière explicite (8.20). Plus précisément si $(W_t^3)_{t \in \mathbb{R}^+}$ désigne le mouvement brownien défini par

$$W_t^3 = (\sqrt{C-1})^{-1} [DW_t^1 + \sqrt{C-1-D^2} W_t^2],$$

alors σ_t^2 est l'unique solution de l'EDS

$$d\sigma_t^2 = (\omega - \theta\sigma_t^2)dt + \sigma_t^2 \sqrt{\alpha(C-1)} dW_t^3. \quad (8.34)$$

Comme $x_0 \geq 0$, σ_t^2 reste bien positif et le processus Y_t s'exprime aisément à l'aide du processus $\sigma_s = \sqrt{\sigma_s^2}$ et du brownien W^1 :

$$Y_t = y_0 + \int_0^t f(\sigma_s^2) ds + \int_0^t \sigma_s dW_s^1;$$

l'intégrale stochastique $\int_0^t \sigma_s dW_s^1$ est bien définie car $E[\int_0^t \sigma_s^2 ds] < \infty$; ce dernier terme se calcul facilement à l'aide de (8.20).

Remarque

Il peut apparaître surprenant que dans le processus limite interviennent deux aléas, alors qu'il n'y en avait qu'un dans les modèles de départ. Ceci est dû au fait que pour mettre en évidence des diffusions limites, nous avons été amenés à étudier des modèles markoviens et que dans le cadre GARCH(1,1) considéré ici, le modèle naturel est le couple markovien (Y_t, σ_{t+1}^2) et par suite le modèle limite est une diffusion bivariée. On peut en outre vérifier que cette diffusion bivariée dépend d'un seul aléa uniquement dans le cas dégénéré où la variable $z_1^{(h)}$ ne prend que deux valeurs.

Remarquons de plus que si $D = E((z_1^{(h)})^3) = 0$, par exemple si $z_1^{(h)}$ a une loi symétrique, le carré de la volatilité σ_t^2 ne dépend que du Brownien W^2 et est indépendant du Brownien W^1 qui régit Y_t ; ceci peut paraître d'autant plus étonnant que dans le modèle GARCH-M, la variance conditionnelle est fonction du passé de Y_t . C'est en fait la non corrélation de $z_1^{(h)}$ et $(z_1^{(h)})^2$ qui entraîne ce phénomène.

Si nous discrétisons le modèle limite ci-dessus par la méthode d'Euler, nous sommes loin du modèle ARCH. Une discrétisation possible qui permettrait de ne faire apparaître qu'un aléa pourrait être, dans le cas $D = 0$, la suivante :

$$\begin{cases} Y_{(k+1)h} - Y_{kh} = f(\sigma_{kh}^2)h + \sigma_{kh}\sqrt{h} z_{k+1} \\ \sigma_{(k+1)h}^2 = (\omega - \theta\sigma_{kh}^2)h + \sigma_{kh}^2\sqrt{\alpha(C-1)}\sqrt{h} \frac{[|z_{k+1}| - (2/\pi)^{1/2}]}{(1 - 2/\pi)^{1/2}} \end{cases}$$

avec z_k , v.a. i.i.d. gaussiennes centrées de variance 1. En effet les deux variables aléatoires :

- z_k et
- $\frac{|z_k| - (2/\pi)^{1/2}}{(1 - 2/\pi)^{1/2}}$

sont centrées réduites et *non corrélées*. Le problème dans cette approximation est que rien n'assure que $\sigma_{(k+1)h}^2$ reste positif.

Pour pallier à cette difficulté, on peut chercher à *discrétiser* le processus $(Y_t, \log \sigma_t^2)$. Ce processus satisfait, si D est quelconque :

$$\begin{cases} dY_t = f(\sigma_t^2)dt + \sigma_t dW_t^1 \\ d(\ln \sigma_t^2) = \left(\omega \exp(-\ln \sigma_t^2) - \theta - \frac{\alpha(C-1)}{2} \right) dt \\ \quad + \sqrt{\alpha} D dW_t^1 + \sqrt{\alpha(C-1-D^2)} dW_t^2 \end{cases}$$

Posons $\alpha_1 = \sqrt{\alpha} D$, $\alpha_2 = \sqrt{C-1-D^2}$ et $\gamma = \theta + \frac{\alpha_1^2 + \alpha_2^2}{2}$.

La discrétisation d'Euler est alors

$$\begin{cases} Y_{(k+1)h} - Y_{kh} = f(\sigma_{kh}^2)h + \sigma_{kh}\sqrt{h} z_{k+1} \\ \ln \sigma_{(k+1)h}^2 - \ln \sigma_{kh}^2 = \left(\frac{\omega}{\sigma_{kh}^2} - \gamma \right) h + \alpha_1\sqrt{h} z_{k+1} \\ \quad + \alpha_2\sqrt{h} \frac{[|z_{k+1}| - (2/\pi)^{1/2}]}{(1 - 2/\pi)^{1/2}} \end{cases}$$

A la différence du modèle GARCH qui n'intègre pas dans la variance conditionnelle le fait que la perturbation σ_{hk} soit de signe positif ou négatif

puisque σ_k n'intervient que par z_k^2 , ce modèle tient compte par la présence de z_k et $|z_k|$ *non seulement de la taille de la perturbation mais aussi de son signe*. Or c'est un fait reconnu que pour certaines séries financières, une hausse de rendement a sur la volatilité un impact différent de celui d'une baisse.

Dans cette optique D.B. Nelson a proposé [1990a] le modèle ARCH - exponentiel autorégressif d'ordre 1 suivant.

8.5.2 Modèle ARCH exponentiel autorégressif d'ordre 1

Dans le dernier modèle présenté ci-dessus, le fait que $\ln \sigma_{(k+1)h}^2$ dépende de $1/\sigma_{kh}^2$ est fort peu agréable, et le modèle proposé par D.B. Nelson fait dépendre $\ln \sigma_{(k+1)h}^2$ de $\ln \sigma_{kh}^2$ (de manière autorégressive) et aussi de z_k et $|z_k|$.

Plus précisément ce modèle est défini de la manière suivante :

$$\begin{cases} Y_t = Y_{t-1} + f(\sigma_t^2) + \sigma_t z_t \\ \ln \sigma_{t+1}^2 = \omega + \beta \ln \sigma_t^2 + \alpha z_t + \alpha' \left(|z_t| - \sqrt{\frac{2}{\pi}} \right) \end{cases} \quad (8.35)$$

où z_t est une suite de v.a. i.i.d. gaussiennes centrées réduites, et où $\omega, \beta, \alpha, \alpha'$ sont des constantes réelles.

Le processus $\ln \sigma_t^2$ est AR(1) et il admet une solution stationnaire si et seulement si $|\beta| < 1$.

Considérons la famille de modèles

$$\begin{cases} Y_{(k+1)h} - Y_{kh} = f(\sigma_{kh}^2)h + \sigma_{kh}\sqrt{h} z_{k+1}^{(h)} \\ \ln \sigma_{(k+1)h}^2 = \omega_h + \beta_h \ln \sigma_{kh}^2 + \alpha_h z_{k+1}^{(h)} + \alpha'_h \left[|z_{k+1}^{(h)}| - \sqrt{\frac{2}{\pi}} \right] \end{cases}$$

et étudions les modèles limites possibles. Remarquons que

$$\ln \sigma_h^2 - \ln \sigma_0^2 = \omega_h + (\beta_h - 1) \ln \sigma_0^2 + \alpha_h z_1^{(h)} + \alpha'_h \left[|z_1^{(h)}| - \sqrt{\frac{2}{\pi}} \right],$$

d'où

$$h^{-1} E[(\ln \sigma_h^2 - \ln \sigma_0^2) | \ln \sigma_0^2 = x] = h^{-1} \omega_h + h^{-1} (\beta_h - 1)x$$

$$h^{-1} \text{var}[(\ln \sigma_h^2 - \ln \sigma_0^2) | \ln \sigma_0^2 = x] = h^{-1} \left[\alpha_h^2 + \alpha_h'^2 \left(1 - \frac{2}{\pi} \right) \right]$$

$$h^{-1} \text{cov}((Y_h - Y_0) (\ln \sigma_h^2 - \ln \sigma_0^2) | \ln \sigma_0^2 = x) = e^{x/2} h^{-1/2} \alpha_h.$$

Nous pouvons alors en déduire la proposition suivante :

Proposition 8.3

Si

$$\begin{cases} h^{-1}\omega_h \rightarrow \omega, & h^{-1}(\beta_h - 1) \rightarrow -\theta \\ h^{-1}\alpha_h^2 \rightarrow \alpha, & h^{-1}\alpha_h'^2 \left(1 - \frac{2}{\pi}\right) \rightarrow \alpha', \end{cases}$$

la famille de chaînes de Markov $(Y_{kh}, \ln \sigma_{kh}^2)$ définie en 8.5.2 et valant (y_0, x_0) à l'instant 0 converge, lorsque h tend vers 0, vers la diffusion bivariable

$$\begin{cases} dY_t &= f(\sigma_t^2)dt + \sigma_t dW_t^1 \\ d \ln \sigma_t^2 &= (\omega - \theta \ln \sigma_t^2)dt + \sqrt{\alpha} dW_t^1 + \sqrt{\alpha'} dW_t^2 \\ (Y_0, \ln \sigma_0^2) &= (y_0, x_0) \end{cases} \quad (8.36)$$

où W^1 et W^2 désignent deux browniens indépendants.

Remarque

Là encore, le modèle limite comporte deux aléas mais, dans ce cas, le processus $\ln \sigma_t^2$ est un processus d'Ornstein-Uhlenbeck. Nous savons que ce processus admet une solution stationnaire si et seulement $\theta > 0$. Or le modèle discret σ_{kh}^2 , processus AR (1), admet une solution stationnaire si et seulement si $|\beta_h| < 1$. L'ensemble est cohérent car $\theta = \lim_{h \rightarrow 0} \frac{1-\beta_h}{h}$. Nous verrons au paragraphe (8.6) qu'il y aussi convergence des lois stationnaires.

8.5.3 Modèle GARCH à seuil (M. El Babsiri et J.M. Zakoian [1990])

Ce modèle tient compte comme l'exemple précédent d'une asymétrie dans la réponse de la volatilité aux variations du rendement. Il s'écrit

$$\begin{cases} Y_t &= Y_{t-1} + f(\sigma_t^2) + \sigma_t z_t \\ \sigma_{t+1} &= \omega + \beta \sigma_t + \alpha^+(\sigma_t z_t)^+ - \alpha^-(\sigma_t z_t)^- \end{cases} \quad (8.37)$$

où z_t est une suite v.a. i.i.d. centrées réduites et où $\omega > 0, \beta \geq 0, \alpha^+ \geq 0, \alpha^- \geq 0$.

Lorsque $\alpha^+ = \alpha^-$, le modèle s'écrit

$$\sigma_{t+1} = \omega + \beta \sigma_t + \alpha \sigma_t |z_t|$$

et a une formulation analogue au modèle GARCH(1,1), le couple $(\sigma_t, |z_t|)$ jouant le rôle de (σ_t^2, z_t^2) .

Dans ce cadre, le modèle de diffusion limite obtenue (cf. El Babsiri, Zakoian [1990]) est de la forme

$$\begin{cases} dY_t = f(\sigma_t^2)dt + \sigma_t dW_t^1 \\ d\sigma_t = (\omega - \theta\sigma_t)dt + \alpha_1\sigma_t dW_t^1 + \alpha_2\sigma_t dW_t^2 \end{cases} \quad (8.38)$$

Le processus σ_t est donc une *diffusion linéaire*, et lorsque les modèles de départ sont symétriques ($\alpha^+ = \alpha^-$), alors $\alpha_1 = 0$ et le processus σ_t dépend seulement du brownien W^2 qui est indépendant du brownien W^1 régissant Y .

Les modèles proposés pour la *volatilité* σ_t dans les divers exemples ci-dessus possèdent des solutions stationnaires si les paramètres sont bien choisis ; ceci sous-entend que la volatilité oscille « de manière stationnaire » autour d'une valeur moyenne, ce qui paraît raisonnable.

Parmi les premiers modèles à *volatilité aléatoire en temps continu* proposés pour des séries financières, on peut citer les modèles proposés par L.O. Scott [1987] :

$$d\sigma_t^2 = \sigma_t^2(\mu dt + \alpha dW_t)$$

par J. Hull et A. White [1987] :

$$d\sigma_t = (\omega - \theta\sigma_t)dt + \alpha dW_t.$$

Le premier présente l'inconvénient de tendre (si $\mu - \frac{\alpha^2}{2} \neq 0$) à vitesse exponentielle soit vers 0, soit vers $+\infty$, lorsque $t \rightarrow +\infty$, quant au deuxième, il possède une solution stationnaire, mais permet à la volatilité de prendre des valeurs négatives.

8.6 CONVERGENCE DES LOIS STATIONNAIRES

Nous allons étudier le modèle GARCH - M, mais des résultats analogues peuvent être obtenus pour les autres exemples.

Considérons le modèle défini en (8.29)

$$\begin{cases} Y_t = Y_{t-1} + f(\sigma_t^2) + \sigma_t z_t \\ \sigma_{t+1}^2 = \omega + \sigma_t^2(\beta + \alpha z_t^2). \end{cases}$$

Nous savons (chapitre 3) que le modèle GARCH (1,1) $\varepsilon_t = \sigma_t z_t$ admet une solution stationnaire stricte si et seulement si $E[\ln(\alpha z_1^2 + \beta)] < 0$ et que la solution peut s'écrire à l'aide d'une série sous la forme suivante :

$$\begin{cases} \varepsilon_t = \sigma_t z_t \\ \sigma_t^2 = \omega + \sum_{i=0}^{\infty} \left(\prod_{k=1}^i (\beta + \alpha z_{t-k}^2) \omega \right). \end{cases}$$

Il est malheureusement impossible en général d'expliciter la loi de cette solution stationnaire.

Ce qu'il est par contre possible de déterminer, c'est le comportement de la queue de la solution stationnaire. Il découle en effet de résultats de H. Kesten [1973] et de Y. Guivarch [1991] sur les solutions d'équations autorégressives à coefficients aléatoires que l'on peut lire le comportement de la queue de la loi de la solution stationnaire σ_t^2 sur la transformée de Laplace suivante :

$$F(\lambda) = E \left[(\alpha z_1^2 + \beta)^\lambda \right].$$

Cette fonction $F(\lambda)$ est bien définie si z_1 a des moments de tout ordre. Elle est strictement convexe et tend vers $+\infty$ lorsque $\lambda \rightarrow +\infty$ si le support de la loi de z_1 est \mathbb{R} .

Supposons que $E[\ln(\alpha z_1^2 + \beta)]$ soit < 0 , alors $F'(0) < 0$ et en 0 la tangente au graphe de F à une pente négative. Il en résulte qu'il existe $\lambda_0 > 0$ tel que $F(\lambda_0) = 1$.

Les auteurs ci-dessus ont alors montré qu'il existe (et c'est un résultat non trivial) une constante $c > 0$ telle que

$$P[\sigma_t^2 > x] \sim \frac{c}{x^{\lambda_0}} \text{ quand } x \rightarrow +\infty. \tag{8.39}$$

Par suite σ_t^2 a des moments d'ordre strictement inférieur à λ_0 et n'a pas de moments d'ordre supérieur ou égal à λ_0 .

On retrouve bien le fait que σ_t^2 a un moment d'ordre 1 si et seulement si

$$F(1) = \alpha + \beta < 1.$$

Dans la mesure où on ne peut connaître la loi stationnaire des modèles GARCH (1,1), il est intéressant de savoir si lorsque la famille de modèles GARCH-M définis en (8.30) converge vers une diffusion limite, la loi stationnaire du processus σ_{kh}^2 converge vers la loi stationnaire de la diffusion limite, loi connue explicitement puisqu'image par l'application $x \rightarrow x^{-1}$ d'une loi Gamma. En effet cette loi limite pourra alors être considérée comme une approximation de la loi invariante obtenue pour le modèle GARCH.

Proposition 8.4

Considérons la famille de chaînes de Markov (Y_{hk}, σ_{kh}^2) définies en (8.30); supposons que lorsque h tend vers 0

- (i) $h^{-1}\omega_h \rightarrow \omega > 0$
- (ii) $h^{-1}(\beta_h + \alpha_h - 1) \rightarrow -\theta$
- (iii) $h^{-1}\alpha_h^2 \rightarrow \alpha$.

Si $\theta > 0$, alors la suite des probabilités invariantes des chaînes de Markov σ_{kh}^2 converge étroitement vers la probabilité invariante de la diffusion limite (8.34)

$$d\sigma_t^2 = (\omega - \theta\sigma_t^2) dt + \sigma_t^2 \sqrt{\alpha(C-1)} dW_t$$

où $C = E((z_1^{(h)})^3)$.

Cette probabilité invariante est l'image par l'application $x \rightarrow x^{-1}$ de la loi Gamma

$$\Gamma\left(\frac{2\omega}{\alpha(C-1)}, 1 + \frac{2\theta}{\alpha(C-1)}\right).$$

La preuve de cette convergence des probabilités invariantes découle du fait qu'à la fois les chaînes de Markov et la diffusion limite issues de n'importe quel point se rapprochent des processus stationnaires à une vitesse exponentielle que l'on peut contrôler.

Remarque

La loi stationnaire de la diffusion limite admet un moment d'ordre 1 si et seulement $\theta > 0$. Il y a en fait les correspondances suivantes entre les modèles GARCH et les diffusions limites :

Modèle	GARCH(1,1)	Diffusion limite
Existence d'une solution stationnaire	$E[\text{Log}(\alpha_h + \beta_h z^2)] < 0$	$\frac{-2\theta}{\alpha(C-1)} < 1$
Moment d'ordre 1	$\alpha_h + \beta_h < 1$	$\theta > 0$

Preuve de la proposition

Soit la famille de chaînes de Markov σ_{kh}^2 définies par

$$\sigma_{kh}^2 = \omega_h + \sigma_{(k-1)h}^2 \left(\beta_h + \alpha_h (z_k^{(h)})^2 \right)$$

où les v.a. $z_k^{(h)}$ sont, pour $k \in \mathbb{N}$, i.i.d. centrées de variance 1.

Nous noterons $\sigma_{kh,x}^2$ la chaîne de Markov issue de $x (\sigma_{0,x}^2 = x)$.

Comme θ est strictement positif, $\alpha_h + \beta_h < 1$ pour h suffisamment petit et le modèle ci-dessus admet une solution stationnaire σ_{kh}^2 de loi ν_h . Remarquons alors que

$$\sigma_{kh,0}^2 - \overline{\sigma_{kh}^2} = \left(\prod_{i=1}^k (\beta_h + \alpha_h (z_i^{(h)})^2) \right) (x - \overline{\sigma_0^2})$$

et par suite

$$E \left(\left| \sigma_{kh,x}^2 - \overline{\sigma_{kh}^2} \right| \right) \leq (\beta_h + \alpha_h)^k E \left(\left| x - \overline{\sigma_0^2} \right| \right).$$

Prenons $x = 0$; alors comme $E(\overline{\sigma_0^2}) = \frac{\omega_h}{1 - \beta_h - \alpha_n}$ (cf.(8.32)), on obtient

$$E \left[\left| \sigma_{kh,0}^2 - \overline{\sigma_{kh}^2} \right| \right] \leq (\beta_h + \alpha_h)^k \frac{\omega_h}{1 - \beta_h - \alpha_h}.$$

D'après le paragraphe 8.5.1, les lois fini-dimensionnelles du processus $(\sigma_t^{(h)})^2$ défini par

$$(\sigma_t^{(h)})^2 = \sigma_{kh,0}^2 \quad \text{si } kh \leq t < (k+1)h,$$

c'est à dire par $(\sigma_t^{(h)})^2 = \sigma_{[t/h]h,0}^2$ convergent vers celles du processus solution de l'EDS (8.34) :

$$\begin{cases} d\sigma_{t,0}^2 = (\omega - \theta\sigma_{t,0}^2) dt + \sigma_{t,0}^2 \sqrt{\alpha(C-1)} dW_t^3. \\ \sigma_{0,0}^2 = 0. \end{cases}$$

L'EDS ci-dessus admet, puisque $\theta > 0$, une solution stationnaire $\overline{\sigma_t^2}$ de loi ν ayant un moment 1. Remarquons qu'il est facile de mesurer l'écart entre $\overline{\sigma_t^2}$ et $\sigma_{t,0}^2$; en effet d'après (8.20)

$$\overline{\sigma_t^2} - \sigma_{t,0}^2 = \xi(t)\overline{\sigma_0^2}$$

où

$$\xi(t) = \exp \left(- \left(\theta + \frac{1}{2} \alpha(C-1) \right) t + \sqrt{\alpha(C-1)} W_t^3 \right)$$

Le processus $\overline{\sigma_t^2}$ ayant un moment d'ordre un, on obtient

$$E \left(\left| \overline{\sigma_t^2} - \sigma_{t,0}^2 \right| \right) \leq \exp(-\theta t) \left(\overline{\sigma_0^2} \right).$$

Il s'agit de prouver que ν_h converge étroitement vers ν lorsque h tend vers 0 de manière à ce que (i), (ii), (iii) soient satisfaites. Il suffit de montrer que pour toute fonction f lipschitzienne bornée, $\nu_h(f)$ converge vers $\nu(f)$. Soit donc f bornée lipschitzienne de rapport K .

Alors, pour tout t réel positif,

$$\begin{aligned}
 |v_h(f) - v(f)| &= \left| \mathbb{E} \left(f \left(\overline{\sigma_{[t/h]h}^2} \right) \right) - \mathbb{E} \left(f \left(\overline{\sigma_t^2} \right) \right) \right| \\
 &\leq K \left(\left| \overline{\sigma_{[t/h]h}^2} - \sigma_{[t/h]h,0}^2 \right| \right) + \mathbb{E} \left(\left| \sigma_{t,0}^2 - \overline{\sigma_t^2} \right| \right) \\
 &\quad + \left| \mathbb{E} \left(f \left(\sigma_{[t/h]h,0}^2 \right) \right) - \mathbb{E} \left(f \left(\sigma_{t,0}^2 \right) \right) \right| \\
 &\leq K \left((\beta_h + \alpha_h)^{[t/h]} \frac{\omega_h}{1 - \beta_h - \alpha_h} + \exp(-\theta t) \mathbb{E} \left(\overline{\sigma_0^2} \right) \right) \\
 &\quad + \left| \mathbb{E} \left(f \left(\sigma_{[t/h]h,0}^2 \right) \right) - \mathbb{E} \left(f \left(\sigma_{t,0}^2 \right) \right) \right|.
 \end{aligned}$$

Le dernier terme de l'inégalité tend, à t fixé, vers 0 lorsque h tend vers 0. On obtient alors à l'aide de (i) et (ii),

$$\limsup_{h \rightarrow 0} |v_h(f) - v(f)| \leq K \exp(-\theta t) \left(\frac{\omega}{\theta} + \mathbb{E} \left(\overline{\sigma_0^2} \right) \right).$$

Il suffit de faire tendre t vers ∞ pour conclure.

8.7 CONCLUSION

Terminons ce chapitre par une comparaison entre les modèles discrets de type hétéroscédastique et les diffusions limites :

En ce qui concerne l'estimation, nous avons le plus souvent à notre disposition une série d'observations discrétisées (par exemple du cours d'un actif) et l'estimation des paramètres d'un modèle GARCH est relativement aisée et en fait plus facile que ceux de la diffusion limite.

En effet dans le cadre GARCH, un *seul aléa* intervient pour décrire le cours de l'actif et sa volatilité et même si la volatilité n'est pas observable, on peut construire (cf. chapitre 4) des fonctions de pseudo-vraisemblance, basées sur l'unique aléa, qui définissent des estimateurs convergents et efficaces.

Pour les modèles de diffusion limites, la volatilité est une variable cachée qui est régie dans certains cas par un brownien indépendant de celui qui intervient dans le cours de l'actif. Une bonne estimation des paramètres définissant la volatilité est encore un problème ouvert.

Au niveau *prévision*, les modèles autorégressifs à erreur GARCH(p, q) sont très bien adaptés et offrent une panoplie de modèles de tailles différentes (dépendant de p et de q). Les modèles limites se limitant au cadre markovien sont beaucoup plus pauvres.

En ce qui concerne les lois stationnaires des divers modèles, le calcul explicite est impossible dans le cadre GARCH(1,1) et est par contre possible pour la diffusion limite. Ceci ajouté au fait que le calcul stochastique permet

souvent des calculs élégants fut une bonne motivation pour considérer en finance les modèles à temps continu émanant des processus hétéroscédastiques. En effet on avait mis en évidence que les modèles autorégressifs à erreurs GARCH étaient bien adaptés aux séries financières, tenant compte d'une *volatilité non constante*. Mais pour tout ce qui est évaluation, par exemple d'*options*, les calculs explicites étaient difficiles (J.-C. Duan [1990]). Or une des raisons de la célébrité du modèle de Black and Sholes pour le prix d'options était la mise en évidence d'un calcul explicite simple. Les modèles à temps continu à volatilité aléatoire ont donc été attrayants et un certain nombre d'articles sur ce thème sont parus (J. Hull et A. White [1987], L.O. Scott [1987], J. B. Wiggins [1987], E. M. Stein et J. C. Stein [1991], E. Renault et N. Touzi [1992], H. Föllmer et M. Schweizer [1993]). Mais en fait les calculs de prix ne sont pas si simples et réclament souvent des simulations de type Monte-Carlo.

Signalons de plus que lorsqu'on modélise le cours d'un actif par un modèle hétéroscédastique discret ou par un modèle à temps continu à volatilité aléatoire, le marché est incomplet et donc trouver la bonne probabilité «risque-neutre» est difficile.

CHAPITRE 9

DE L'UTILITÉ DES MODÈLES ARCH

par

Jean-François BOULIER

Toute nouvelle technique apporte un regard neuf sur une réalité déjà observée et ravive l'espoir de résoudre les questions les plus difficiles ou du moins d'y répondre de façon plus satisfaisante. Les modèles économétriques ARCH n'échapperont pas au cycle d'appropriation humaine qui de la curiosité passe à l'engouement, et de l'excès peut conduire à l'overdose, avant que l'expérience soit suffisante pour que les bons manuels en consacrent l'usage dans un domaine bien balisé. Or nous sommes aujourd'hui en début de cycle.

Apparue et développée pendant les années 80, cette nouvelle technique économétrique a percé sur le constat des imperfections de la modélisation des cours boursiers et des taux au moyen de processus browniens. D'une part les lois marginales des variations relatives des cours ne paraissent pas gaussiennes en raison de leur leptokurtivité, d'autre part l'autocorrélation des carrés ou des cubes des variations relatives des cours est significativement positive. Extension assez proche du modèle de base d'un grand nombre de techniques d'évaluation, les modèles ARCH sont très séduisants : ils constituent un pas dans la bonne direction tout en restant dans un domaine de complexité abordable.

Pour ces raisons, l'équipe de recherche innovation du CCF a depuis 1988 mené de nombreuses recherches sur les modèles ARCH. Nous nous y sommes tout d'abord intéressés à des fins de description et de mise en œuvre de procédure d'estimation, nous avons ensuite cherché à utiliser cette approche en vue d'améliorer nos prévisions, en envisageant notamment la prévision de volatilité. L'évaluation des options et la détection d'arbitrage a fait l'objet de quelques tentatives. Parallèlement, l'intérêt potentiel de cette modélisation économétrique dans le con-

texte de la gestion de portefeuille a été plusieurs fois jaugé. Le fil conducteur est, on l'aura compris, la modélisation du risque et ses applications.

Il est clairement trop tôt pour porter un jugement définitif car le manque de recul et d'expérimentation pourrait nous laisser ignorer leur utilité ou nous méprendre sur leur efficacité. Le but de cet exposé est donc de rassembler quelques illustrations des quatre étapes de nos recherches, dont certaines ont fait l'objet de communications spécifiques, et de faire le point sur l'utilité potentielle pour l'industrie financière de cette intéressante modélisation. Il s'agira pour l'essentiel de résultats empiriques. Le lecteur ne s'étonnera donc pas d'y trouver des conclusions moins positives et plus partielles que dans la littérature sur le sujet. On se rapportera aux précédents chapitres pour trouver les références de base, les descriptions des modèles et leur estimation. Le point de vue retenu ici est celui d'un utilisateur, dont on ne peut prétendre qu'il soit tout à fait représentatif, mais dont l'ambition est de tirer le meilleur profit de la modélisation avec un esprit d'enthousiasme critique.

9.1 POUVOIR DESCRIPTIF

Nous nous intéressons au comportement statistique de séries de rentabilités r_t d'actifs financiers, incorporant les coupons ou dividendes s'il y a lieu. La modélisation désormais classique qui est à la base de l'évaluation d'option ou des choix de portefeuille peut se résumer à :

$$r_t = R\Delta t + \sigma Z_t \quad (9.1)$$

où R est la rentabilité espérée et σ la volatilité ;

t est l'intervalle de temps ;

Z_t est une variable aléatoire gaussienne d'espérance limitée et de variance Δt .

Cette modélisation est bonne dans la mesure où elle reconnaît le caractère aléatoire des variations de prix. Mais elle trébuche dans le détail sur la représentation statistique. A titre d'exemple, les graphiques suivants extraits de Rabemanajara et Batt [1991] illustrent les distributions des coefficients d'asymétrie (skewness, figure 9.1) et d'aplatissement (kurtosis, figure 9.2) d'un échantillon de 113 valeurs françaises (données hebdomadaires 81-89). Sachant que ces coefficients devraient être nuls dans le cas d'une loi normale, ce que l'on rejette au seuil de 5% lorsqu'ils sont supérieurs à 0.045 et 0.091 respectivement, il est visible que la modélisation n'est pas très représentative de la réalité boursière. Cette conclusion est un peu nuancée lorsque l'on change la fréquence des observations. On remarque sur les graphiques qui présentent les distributions empiriques du CAC 240 (figure 9.3) que l'aplatissement est surtout prononcé pour les cours quotidiens et hebdomadaires et qu'il a tendance à s'estomper avec une fréquence mensuelle.

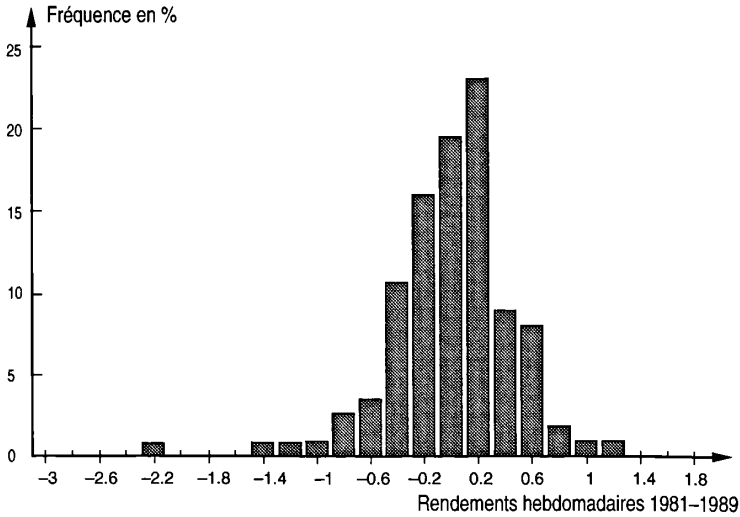


Figure 9.1: Distribution des skewness des valeurs

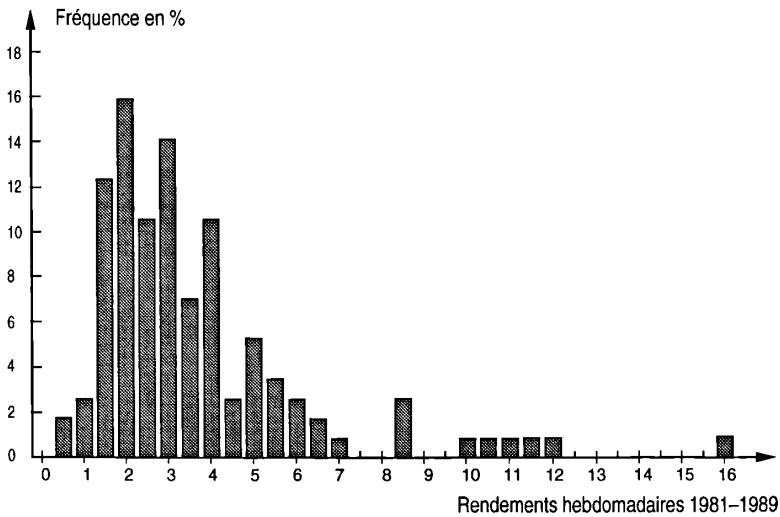


Figure 9.2: Distribution des kurtosis des valeurs

A partir de la constatation de l'inadéquation du modèle (9.1), au moins deux voies se dégagent : ou bien changer la loi de probabilité de Z ou bien relâcher l'hypothèse d'une volatilité constante. C'est ce dernier choix que nous allons faire dans cet exposé en essayant de discerner les avantages et les limites de cette nouvelle modélisation.

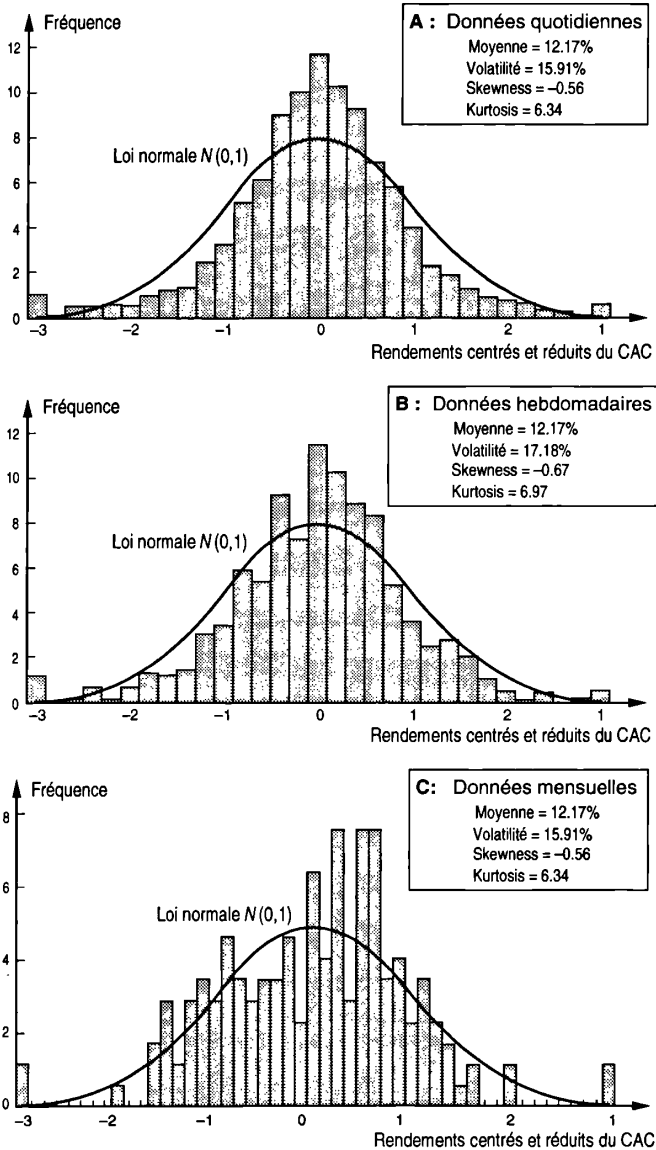


Figure 9.3

Distribution d'une loi normale et distribution empirique de l'indice CAC pour des données quotidiennes, hebdomadaires et mensuelles (1976-1990)

9.1.1 Modélisation GARCH de l'indice CAC

Afin de voir dans quelle mesure ce type de processus est adapté à la modélisation des rendements d'un actif financier, nous avons estimé (voir Rabe-manajara et Batt [1991]) sur les rendements quotidiens du CAC les paramètres d'un modèle GARCH(1,1) sur la période 1976-1990, en introduisant dans la moyenne un terme autorégressif d'ordre 1, en raison de l'existence d'une autocorrélation entre les rendements quotidiens sur cette période. Le processus autorégressif à erreurs GARCH(1,1) représentant $Y_t = \log(1 + r_t)$ est le suivant (l'écart-type d'estimation figure entre parenthèses sous la valeur estimée du paramètre) :

$$Y_t = 0.00056 + 0.26Y_{t-1} + e_t \quad (9.2)$$

(0.0001) (0.017)

$$V(e_t | I_{t-1}) = \sigma_t^2 = 0.042 + 0.156e_{t-1}^2 + 0.807\sigma_{t-1}^2 .$$

(0.0045) (0.009) (0.011)

Les rendements quotidiens de l'indice CAC en 1987 ainsi que les volatilités calculées à chaque date à l'aide du modèle estimé ci-dessus sont présentés sur le graphique de la figure 9.4 : si la volatilité estimée par le modèle avant le krach boursier d'octobre 1987 était généralement inférieure à 20%, elle a atteint d'après le modèle plus de 60% au cours du dernier trimestre de l'année 1987.

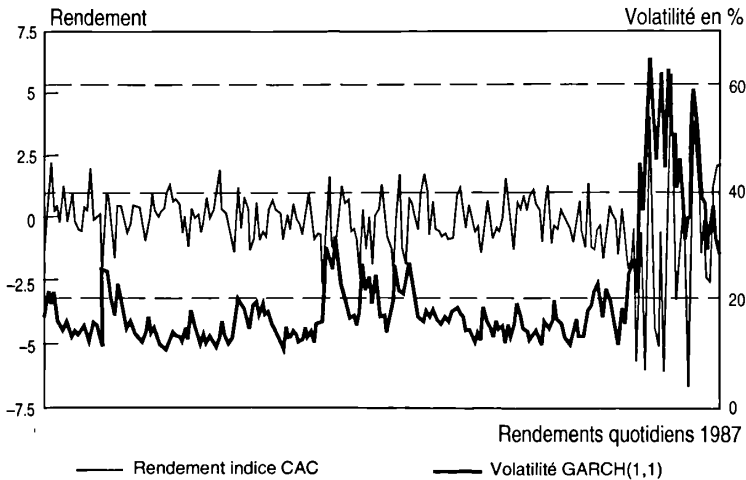


Figure 9.4

Rendements quotidiens de l'indice CAC Général en 1987 et volatilité estimée sous l'hypothèse du modèle GARCH(1,1)

La statistique de test du rapport de vraisemblance du modèle contre le modèle GARCH(1,1) vaut 972 et conduit facilement au rejet de l'hypothèse d'homoscédasticité. Il est intéressant de comparer également les caractéristiques

des distributions empiriques du bruit e_t/σ sous l'hypothèse d'homoscédasticité (variance constante) et celles de e_t/σ_t sous l'hypothèse du modèle GARCH(1,1) estimé ci-dessus. Les distributions empiriques de ces processus sont représentés à la figure 9.5 avec la densité d'une loi gaussienne. On voit que la distribution de e_t/σ_t est plus proche de celle d'une loi normale que celle de e_t/σ . De plus, la modélisation GARCH permet de réduire très significativement la kurtosis qui passe de 6.45 pour e_t/σ à 2.98 pour e_t/σ_t .

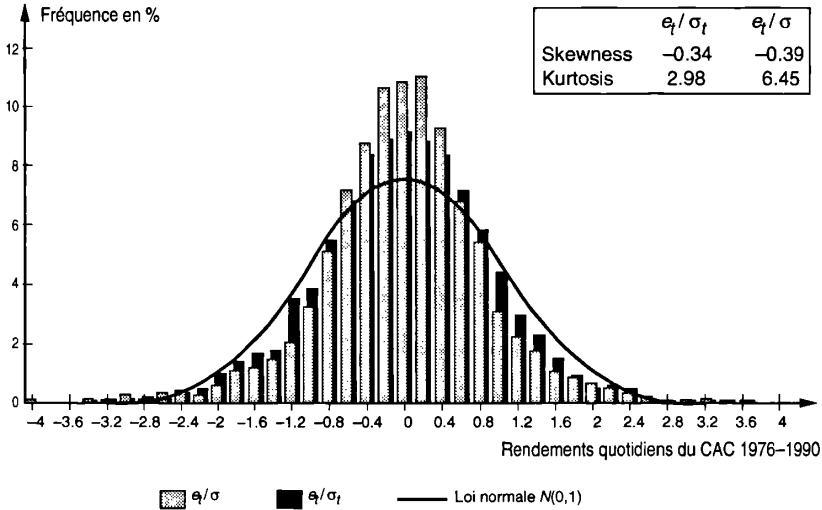


Figure 9.5

Rendements quotidiens de l'indice CAC de 1976 à 1990 : distributions de e_t/σ_t et de e_t/σ

9.1.2 Modélisation GTARCH de l'indice CAC

La modélisation GARCH qui vient d'être présentée rend bien compte des poussées de fièvre du marché financier parisien. Mais les chocs les plus sévères sont plutôt à la baisse. Le sentiment boursier est corroboré par l'observation des autocorrélations des résidus ε_t du modèle AR du CAC sur la période 1976-1991 : il est visible que la baisse entraîne la baisse car l'autocorrélation des $\varepsilon_t^- = \min(0, \varepsilon_t)$ est plus forte que l'autocorrélation des hausses ($\varepsilon_t^+ = \max(0, \varepsilon_t)$), comme le montre la figure 9.6.

Les variations vécues le plus douloureusement ne seraient-elles pas perceptibles, en y incorporant une différenciation de la variance conventionnelle selon la hausse ou la baisse du marché ? C'est l'objectif de la modélisation TARCH à seuil (T : threshold) décrite par l'école française et dont l'application à l'indice CAC est décrite dans Rabemananjara et Zakoian [1991].

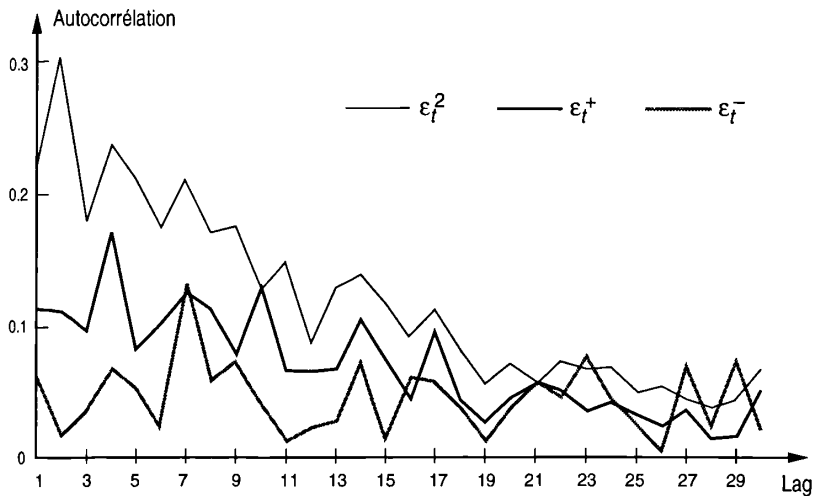


Figure 9.6: Fonctions d'autocorrélation : séries de résidus

Parmi les modèles estimés dans Rabemananjara et Zakoian [1991], le modèle GTARCH non contraint répond aux spécifications suivantes :

$$\begin{aligned} r_t &= c + d_1 r_{t-1} + \varepsilon_t \\ \varepsilon_t &= \sigma_t Z_t \end{aligned} \quad (9.3)$$

où Z_t est le tirage d'une même variable aléatoire symétrique et

$$\sigma_t = \alpha_0 + \alpha_1^+ \varepsilon_t - \alpha_1^- \varepsilon_t^- + \beta_1 \sigma_{t-1}$$

avec $\alpha_0 > 0$ mais sans contraintes sur les autres coefficients.

Les résultats sont répertoriés dans le tableau 9.1 (page suivante).

Le marché français semble donc réagir plus brusquement à la baisse ($\alpha_1^+ < \alpha_1^-$). Des auteurs mentionnent des variances à la hausse significativement plus faibles que celles à la baisse.

Des comportements similaires ont également été identifiés pour plusieurs valeurs de la cote, mais dans plusieurs cas certains coefficients sont négatifs.

Tableau 9.1: Estimation des paramètres 1976-1990

Paramètres	Homoscédastique	GARCH(1,1)	(3)
Constant	0.040 (0.016)	0.056 (0.013)	0.038 (0.013)
Θ_1	0.182 (0.016)	0.262 (0.017)	0.262 (0.017)
α_0	0.989	0.042 (0.007)	0.049 (0.017)
α_1^+ ou α_1^-		0.156 (0.009)	0.111 (0.020) 0.192 (0.019)
β_1		0.807 (0.011)	0.833 (0.020)
Variance historique L-R test d'homoscédasticité	0.989	1.135 486.0	1.269 501.2

9.1.3 Modélisation du MATIF

Ces modèles sont-ils stables ? La réponse est malheureusement souvent négative ; d'ailleurs plus la modélisation est sophistiquée et moins elle est robuste. A titre d'exemple, nous présentons dans le tableau 9.2 les variations de paramètres que nous avons constatées (voir Jamet et Rabemanajara [1989]) dans le cas du Matif, pour lequel ont été étudiés les contrats Juin 1988, Décembre 1988 et Mars 1989.

L'hypothèse d'homoscédasticité ne peut être rejetée pour la deuxième échéance et les coefficients sont assez différents entre la première et la dernière : il est à noter que les volatilités des lois marginales valent respectivement 11.7% pour Juin 1988 et 5,1% pour Mars 1989 signe que le krach de 1987 a très nettement affecté les cotations du premier contrat.

Ce choix de quelques illustrations vise à étayer nos principales conclusions pratiques :

- les modèles ARCH apportent une meilleure description que le modèle classique,
- la gamme des modèles est très vaste et l'intuition peut être un bon guide,
- la stabilité temporelle des paramètres est rarement au rendez-vous.

Tableau 9.2: Estimation des paramètres du modèle GARCH(1,1)

Échéance Période d'estimation	Paramètres estimés ¹				Test ² de l'hypothèse d'homoscédasticité
	<i>m</i>	<i>c</i>	<i>alpha</i>	<i>beta</i>	
Juin – 88 1/11/87-20/5/88	0.02 (0.4)	0.06 (1.8)	0.32 (3.0)	0.57 (3.7)	110.1
Déc-88 5/4/88-18/11/88	0.04 (1.4)	0.08 (0.6)	0.09 (0.6)	0.40 (0.4)	4.18
Mars-89 1/7/88-20/2/89	0.04 (1.1)	0.08 (1.1)	0.19 (1.2)	0.40 (1.0)	43.2

¹ La statistique du test de Student figure entre parenthèses sous le coefficient.
² Sous l'hypothèse d'homoscédasticité, la statistique suit une loi du chi deux à 2 ddl dont la valeur critique au seuil de 5% vaut 6

9.2 PREVOIR AVEC LES ARCH

Les illusions des financiers sont tenaces : on veut croire que les marchés sont (pour partie) prévisibles. Puisque les modèles ARCH paraissent bien représenter la variation des cours, n'y a-t-il pas de l'information exploitable pour prévoir l'évolution future du marché ? La première idée est naturellement d'essayer de prévoir la volatilité (implicite) du marché des options. Les modèles GARCH-M constituent une deuxième piste dans laquelle la prévision multivariée s'enrichit des impacts de la variabilité sur la variable prédite.

9.2.1 Prevoir la volatilité du MATIF

Le marché des options Matif a dès son lancement été très liquide en ce qui concerne les séries d'options à la monnaie. C'est pourquoi, dans Jamet et Rabemananjara [1989], et sur les échéances des contrats futures de Juin et Décembre 1988 ainsi que de Mars 1989 (cf 9.1.3) ont été calculées les volatilités implicites du marché (modèle de Black – option européenne) sur les cotations de clôture.

Les modèles GARCH présentés au 9.1.3 ont alors servi à prévoir la volatilité correspondant à un contrat et une maturité H (exprimée en jours) au moyen de :

$$V_t^2(H) = \sum_{h=1}^H \frac{\sigma_{t+h}^2}{H} \quad (9.4)$$

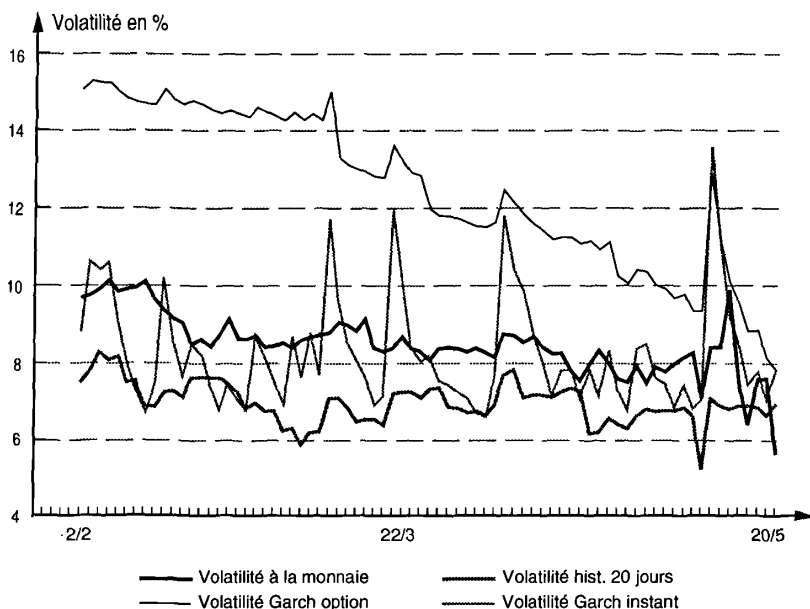
où σ_{t+h} est l'espérance conditionnelle à la date t de la volatilité à la date $t+h$.

Dans le tableau 9.3 sont reportés, en fonction des trois échéances, les coefficients de corrélation de la volatilité implicite avec d'une part la volatilité historique (20 bourses) et V_t .

Tableau 9.3: Coefficients de corrélation de la volatilité implicite

Echéance	Volatilité historique	Volatilité GARCH
Juin-88	0.47	0.82
Déc-88	-0.047	0.58
Mars-89	0.59	0.37

Il semble bien qu'une partie des variations du marché soient anticipables au moyen d'un modèle GARCH, toutefois le succès n'est pas assuré. D'autre part, comme il est visible sur le graphique correspondant à l'échéance Juin 1988, le niveau absolu de la volatilité implicite peut rester (traumatisme du krach) très éloigné du niveau prévu au moyen de la modélisation.

**Figure 9.7:** Volatilités

9.2.2 Prévoir la volatilité des cours de change

La même démarche nous a conduit (voir Eyssartier et Le Douarec [1989]) à analyser les cours d'options du marché de Philadelphie sur les couples GBP/USD et DEM/USD, c'est à dire les cours de change de la livre et du Deutsche Mark en Dollars. Les modélisations GARCH ayant été déterminées, les résultats du

modèle (4) sont représentés sur les graphiques suivants où ils sont comparés à la volatilité historique (60 jours) et la volatilité implicite (modèle de Garman et Kohlhagen). Il est visible que le modèle GARCH fournit de l'information sur ces marchés d'options. Les impacts de l'invasion du Koweït et du déclenchement de la guerre terrestre sont assez manifestes.

9.2.3 Modèles GARCH-M pour prévoir le CAC

Notre ambition (voir Eyssartier et Le Douarec [1989]) est là beaucoup plus grande : est-ce que la modélisation ARCH peut nous permettre de concevoir des stratégies qui battent le marché d'actions ?

Voici une sélection de trois spécimens parmi les nombreux modèles de prévision testés dans Eyssartier et Le Douarec [1989]. Les variables explicatives sont déterminées dans un échantillon de taux de croissance d'indices boursiers mondiaux dont le CAC lui-même, de taux d'intérêts et de valeurs françaises. Nous avons volontairement limité le nombre de variables à trois et les avons déterminées (avec éventuellement des retards) sur la base de la variance expliquée. A partir de là, les erreurs du modèle ont été représentées au moyen d'un modèle GARCH(1,1) ou d'un modèle GARCH-M(1,1).

$$\varepsilon_t = r_t - \sum_{1 \leq i \leq 3} \alpha_i x_{i,t} - \delta \sigma_t \quad (9.5)$$

$$E(\varepsilon_t) = 0$$

$$\text{Var}(\varepsilon_t) = \sigma_t^2$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

Les performances de stratégies d'investissement fondées sur les trois modèles (M1 : régression simple, M2 : régression GARCH, M3 : régression GARCH-M) sont résumées dans le tableau 9.4.

En considérant que notre prévision est la moyenne d'une gaussienne d'écart type déduit de la variance conditionnelle, il est aisé de calculer une probabilité P de hausse du CAC. La stratégie testée consiste à investir toutes les semaines 100 F sur le CAC si P est supérieur à 55% et à le vendre à découvert pour 100 FRF si P est inférieur à 45%. La position est débouclée en fin de semaine. Dans l'intervalle $45\% \leq P \leq 55\%$ on ne prend pas position.

La deuxième colonne du tableau 9.4 indique le nombre de semaines gagnantes (sur 56 semaines de Février 88 à Mars 89) les colonnes 3 à 5 fournissent les erreurs (ε_t) moyennes en valeur absolue, absolues moyennes et quadratiques moyennes de cette stratégie. Enfin la colonne 6 fournit le gain moyen, son écart type étant en colonne 7 et leur rapport appelé ratio de Sharpe, figure dans la dernière colonne.

Tableau 9.4
Performances des stratégies d'investissement fondées sur trois modèles

Modèle	Test du signe	Erreur moyenne (%)	Erreur absolue moyenne (%)	Erreur quadratique moyenne (%)	Gain	Écart type	Ratio de Sharpe
M1	33	0.51	1.63	1.94	28.95	1.64	0.32
M2	33	0,54	1.64	1.93	32.12	1.42	0.41
M3	29	0.96	1.74	1.92	8.53	1.71	0.09

M1 : régression simple ; M2 : régression GARCH ; M3 : régression GARCH-M

Ces résultats obtenus en dehors de la période d'estimation ne sont pas mauvais. Les ratios de Sharpe sont positifs — toutefois, ils ne sont pas exceptionnels car un bon ratio serait plutôt supérieur à 0.5. Le modèle GARCH sort nettement du lot (sa performance est illustrée à la figure 9.8). En revanche le modèle GARCH-M semble contreperformer. L'explication en est sans doute qu'en période de forts mouvements il y a une double inhibition du signal de ce modèle — d'une part la moyenne baisse et d'autre part la variance monte. Comme le CAC a été plutôt orienté à la hausse sur la période, les signaux de M3 ne sont pas très efficaces. Serait-il meilleur en période de baisse ?

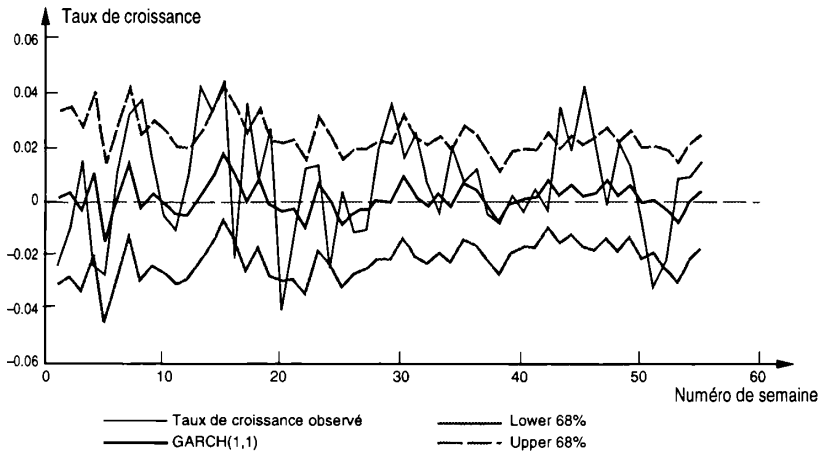


Figure 9.8: Comparaison en prévision des modèles : prévision de l'indice CAC

Un mot de commentaires pour conclure cette partie « prévisions ». Anticiper le marché est une gageure, c'est bien connu... mais le taux de réussite

est finalement parfois alléchant. Les modèles ARCH nous apportent dans ce domaine :

- d'intéressants signaux sur le marché des options, mais dont la fiabilité (stabilité notamment) est incertaine ;
- un signal complémentaire utile sur le résidu d'un modèle de régression linéaire ; là encore trop de sophistication semble nuire.

9.3 LES OPTIONS DANS UN UNIVERS HÉTÉROSCÉDASTIQUE

Les méthodes précédentes (cf 2.1 et 2.2) manquent de cohérence ! Si les volatilités sont variables pourquoi utiliser les signaux « ARCH » pour prévoir la volatilité implicite des modèles qui présupposent des volatilités constantes ! Il paraît utile de dépasser le cadre des hypothèses de Black et Scholes et de tenter d'apprécier l'apport d'une modélisation plus riche. Plusieurs recherches théoriques sont conduites sur le sujet, en particulier en France (voir les communications du Congrès 1992 de l'AFFI). Nos travaux remontent à deux ans. Ils sont rapportés dans Pluchet et Sauvart [1990], et suivent de près la modélisation de Hull et White [1987]. Comme ils n'ont pas fait objet de communication, nous détaillerons les résultats en présentant tout d'abord le cadre de la modélisation puis en évoquant les questions d'évaluation numérique. On montrera que le pricing ne diffère de celui de Black et Scholes que dans le cas des options de maturités courtes et son efficacité sera illustrée dans le cas d'options sur le CAC.

9.3.1 Modèle stochastique et valorisation d'option

Les modèles GARCH s'avèrent intéressants on s'est efforcé de trouver les valeurs d'options dans le cadre d'un modèle stochastique dont la « discrétisation » soit un GARCH(1,1). Les travaux de Nelson (voir Pluchet et Sauvart [1990]) montrent que le modèle discret définit par :

$$n \Delta \log(S_t) = \left(\mu - \frac{\sigma_t^2}{2} \right) \Delta t + \sigma_t Z_t$$

$$\sigma_{t+1}^2 = \omega + \alpha (\sigma_t Z_t)^2 + \beta \sigma_t^2 \quad (9.6)$$

où Z_t suit une loi normale $N(0, \Delta t)$ et Z_t est indépendant de Z_{t-1} , a pour limite le modèle stochastique suivant :

$$\frac{dS_t}{S_t} = \mu dt + \sigma_t dW_1$$

$$d\sigma_t^2 = (\omega - \Theta \sigma_t^2) dt + \alpha \sigma_t^2 dW_2 \quad (9.7)$$

où W_1 et W_2 sont deux mouvements browniens indépendants et $\Theta = 1 - \alpha - \beta$.

Des développements complexes de calcul stochastique s'inspirant de ceux décrits dans Hull et White [1987] qui supposaient que $\omega = 0$, permettent d'obtenir la valeur du call européen d'échéance T par :

$$\text{Call} = \int_0^{\infty} BS(y)g(y)dy \quad (9.8)$$

où

$$y = \int_0^T \sigma_t^2 dt$$

et BS est la valeur du Call donnée par Black et Scholes avec une volatilité y du cours.

9.3.2 Évaluation numérique

Le calcul numérique approché de (9.8) peut être conduit de plusieurs façons :

- en effectuant un développement de Taylor sous le signe intégrale à un ordre suffisant ; les ordres que nous avons choisis sont 1 (méthode n° 1) et 2 (méthode n° 2) qui nécessitent le calcul des moments d'ordre 1 et 2 de y . Le calcul analytique complet de ces moments a pu être mené dans le cadre du modèle (9.7).
- en effectuant des simulations de Monte Carlo, un processus (9.6) peut être généré, puis la distribution empirique de y ainsi obtenue sert à évaluer (9.8) (méthode n° 3). Les résultats de 20 000 tirages montrent que y paraît log-normalement distribué.
- sous l'hypothèse que y soit une loi log-normale (dont on connaît exactement les moments), on peut calculer l'intégrale numériquement (méthode n° 4) avec une très bonne précision (méthode des rectangles par exemple).

Dans le tableau 9.5 sont reportées les valeurs obtenues avec les 4 méthodes pour des options de maturité 100 jours, un taux de 7.9%, une valeur de $S = 98.7$ (donc un prix à terme de 101.75) et des paramètres $\alpha = 0.4$, $\beta = 0.45$, $\omega = 0.186$, en fonction de leur prix d'exercice K .

Comme on le constate, ces valeurs sont très groupées et les écarts relatifs sont inférieurs à 2% — ils sont plus importants bien entendu, pour les options en dehors de la monnaie. Le calcul numérique montre que le coefficient en facteur de $\text{Var}(y)$ est négatif à la monnaie et positif en dehors, ce qui est conforme à l'observation du marché (effet « sourire »). Nous utiliserons le premier modèle dans la suite.

Tableau 9.5

Valeur du Call suivant 4 méthodes de calcul et 3 hypothèses de leur prix d'exercice K

K	Méthode	Call	Put
100	1	2.389	0.6922
	2	2.380	0.6837
	3	2.371	0.6739
	4	2.374	0.6767
102	1	1.271	1.513
	2	1.262	1.505
	3	1.248	1.491
	4	1.250	1.493
104	1	0.5720	2.755
	2	0.5637	2.746
	3	0.5602	2.743
	4	0.5635	2.746

9.3.3 Volatilités effectives

Les résultats numériques précédents indiquent que dans une plage de prix d'exercice, le prix de l'équation (9.8) peut être approché par le $BS(E(y))$, autrement dit que l'on peut en pratique conserver la formule donnée par Black et Scholes à condition d'utiliser une volatilité effective qui correspond à l'espérance de y qui peut être calculée exactement :

$$E(y) = \frac{\omega}{\Theta} T + \frac{k}{\Theta^2} (1 - e^{-T\Theta}) \quad (9.9)$$

avec $k = \sigma_0 \Theta - \omega$.

La volatilité effective vaut donc :

$$V = \sqrt{\frac{\omega}{\Theta} + \frac{k}{\Theta^2} \frac{1 - e^{-T\Theta}}{T}} \quad (9.10)$$

$\sqrt{\omega/\Theta}$ peut-il s'interpréter comme une volatilité de long terme et $1/\Theta$ comme un temps caractéristique ?

Cette fonction $V(T)$ est représentée sur la figure 9.9 dans le cas d'une variance initiale deux fois supérieure à la variance de long terme et avec les paramètres suivants : $\alpha = 0.1$, $\beta = 0.86$ et $\omega = 0.033$. Le graphique montre

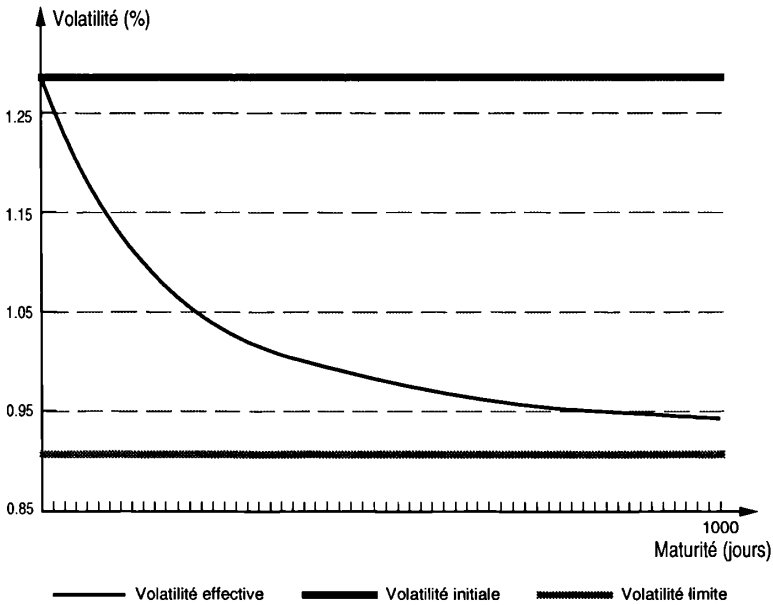


Figure 9.9: Volatilité effective

ainsi qu'au delà d'une centaine de jours, l'écart entre la volatilité effective et la volatilité de long terme est très faible. Cela fournit également une explication au fait empiriquement connu que pour un même prix d'exercice et un même sous-jacent la volatilité du marché puisse dépendre de l'échéance de l'option.

9.3.4 Stratégie d'arbitrage des options CAC

Comme l'écart d'évaluation est surtout sensible lorsque l'échéance de l'option est très rapprochée, nous avons testé s'il était possible d'effectuer des arbitrages sur le marché des options du CAC 40.

10 jours ouvrables avant l'échéance du contrat d'option, on décide de vendre un call et un put à la monnaie si leur valeur de marché est plus élevée que celle donnée par le modèle. On achète un call et un put à la monnaie dans le cas contraire.

Le jour d'échéance, on gagne (respectivement perd) $|S - K|$ si l'on est porté acheteur (respectivement vendeur).

Les résultats obtenus entre novembre 1988 et mars 1990 sont présentés dans le tableau 9.6. La valeur de Call+Put à la monnaie évaluée à la volatilité historique est indiquée pour mémoire.

On ne peut pas toujours trouver des options à la monnaie dans la pratique. Si tel est le cas on achète (ou on vend) pour moitié les options les plus proches de la monnaie de part et d'autre.

Tableau 9.6
Test des stratégies d'arbitrage sur le marché des options du CAC 40

Mois	Garch	Marché	Historique	Décision	Gain
Nov-88	29	48	26	VV	+41
Déc-88	23	29	12	V	-12
Jan-89	34	48	37	V	+11
Fév-89	35	61	34	VV	-9
Mar-89	44	58	38	V	+37
Avr-89	34	53	22	VV	+47
Mai-89	36	44	29	V	-4
Jui-89	33	36	19	V	+28
Jui-89	36	31	33	A	+4
Aoû-89	34	38	19	V	-12
Sep-89	34	44	26	V	+35
Oct-89	70	117	123	VV	+97
Nov-89	46	56	36	V	+1
Déc-89	42	56	36	V	+18
Jan-90	48	62	43	V	+7
Fév-90	46	57	41	V	+50
Mar-90	48	50	48	V	+35

On constate que dans la plupart des cas le marché est trop cher : l'estimateur fondé sur le modèle GARCH et celui fondé sur la volatilité historique à court terme (10 jours) conseillent souvent à juste titre de se porter vendeur. L'estimateur GARCH est plus stable que celui fondé sur la volatilité historique à court terme. Il est possible que ces gains soient attribuables à la phase de lancement du marché que nous avons considérée.

Lors du mini-krach d'octobre 1989, la volatilité implicite s'aligne sur la volatilité historique à 10 jours. La volatilité GARCH augmente aussi, mais moins. En octobre 1989, l'estimateur GARCH s'avère meilleur que la volatilité historique et que le marché puisque l'on peut constater ex post que le support à moins varié que prévu.

Beaucoup de travail attend encore le modélisateur dans le domaine de l'évaluation des options avec volatilité ARCH et d'une façon générale dans tout type de gestion dynamique. En pratique toutefois, sous réserve de tests plus complets, il semble que le bénéfice de ces modélisations ne soit tangible que dans le très court terme (un mois tout au plus).

9.4 GESTION DE PORTEFEUILLE ET MODÉLISATION ARCH

L'analyse du risque constitue le point de départ de la gestion quantitative. Aussi est-il tentant de considérer les améliorations potentielles des techniques modernes de gestion en ce qui concerne les concepts statiques, le bêta, le risque spécifique ou la distribution des rentabilités futures du portefeuilles ou en ce qui concerne la gestion proprement dite, choix d'actifs selon des primes de risque par exemple ou bien l'optimisation.

Ayant fait plusieurs essais dans chacun de ces domaines, nous avons choisi de présenter ceux dont les résultats sont les plus surprenants et qui se rapportent à l'optimisation du portefeuille (voir Boulier et Rabemananjara [1992] pour plus de détails).

9.4.1 Persistance d'effets ARCH

Sur la figure 9.10 sont représentés les coefficients d'applatissage de portefeuilles d'actions françaises en fonction du nombre de lignes les composant (investissements équipondérés voir Rabemananjara et Batt [1991]). Il apparaît que l'applatissage ne disparaît pas. D'ailleurs dans le cas de l'indice de marché — qui est un portefeuille particulier — l'effet ARCH est tout à fait significatif (voir ci-dessus, section 1.1).

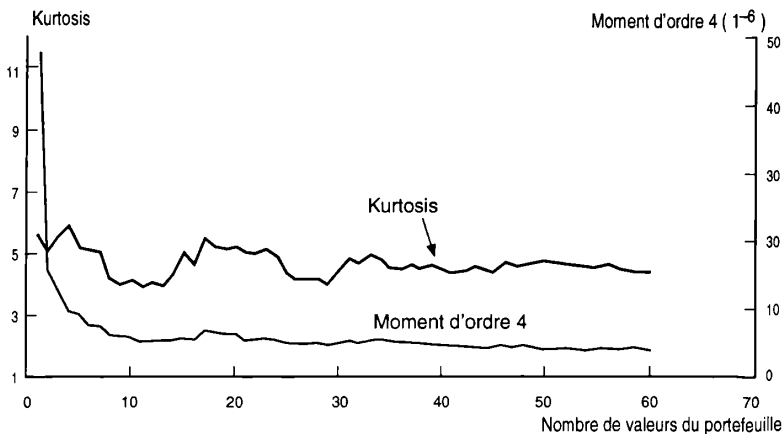


Figure 9.10

Kurtosis d'un portefeuille en fonction du nombre de valeurs: données hebdomadaires 1981–1989; tri par capitalisation boursière

9.4.2 Modèles à facteurs latents hétéroscédastiques

Les questions d'allocations d'actifs sont parmi les plus importantes en gestion de portefeuille. La démarche suivie en gestion quantitative consiste dans

un premier temps à analyser d'une part l'information conduisant à faire des paris et d'autre part à mesurer les risques associés à ces paris. Enfin le portefeuille est construit au moyen d'une optimisation à niveau de risque donné des pondérations dans chaque classe d'actifs, ce que nous effectuerons dans la section suivante.

Le modèle à facteurs latents hétéroscédastiques vise à mieux représenter les risques joints des classes d'actifs dans lesquels on va investir. Prenons le cas des marchés d'actions des cinq principaux marchés mondiaux : force est de constater que leurs turbulences sont généralement concomitantes. Il est donc utile de prolonger la modélisation GARCH de l'indice de marché (cf 1.2) à un modèle représentant globalement les variations des bourses. C'est l'objectif du modèle de Diebold et Nerlove [1989] :

$$\begin{aligned}
 r_t &= C f_t + u_t & (9.11) \\
 (5,1) & \quad (5,k) (k,i) & (5,1) \\
 E_{t-1}(f_t) &= 0 \\
 E_{t-1}(u_t) &= 0 \\
 E_{t-1}(f_{it}u_{jt}) &= 0 \\
 E_{t-1}(f_t f_t') &= \Lambda_t, & \text{matrice diagonale} \\
 E_{t-1}(u_t u_t') &= \Gamma.
 \end{aligned}$$

L'hétéroscédasticité se situe au niveau des éléments diagonaux de la matrice Λ_t . Dans Boulier et Rabemananjara [1992], nous avons utilisé une représentation GARCH(1,1) :

$$\omega_{i,t}^2 = \alpha_0 + \alpha_i f_{i,t-1}^2 + \beta_i \omega_{i,t-1}^2 \tag{9.12}$$

et une représentation TARCH(1,1).

9.4.3 Simulation de portefeuille et performances

Ayant estimé le modèle sur des données antérieures à 1985, nous avons réalisé une comparaison de stratégies de gestion toutes fondées sur la minimisation du risque du portefeuille pour un investisseur français pendant la période 1985-1991, mais en appréciant le risque soit au moyen de la matrice de covariance historique soit au moyen des modélisations GARCH et TARCH décrites dans la section précédente. Nous n'avons retenu qu'un seul facteur latent. La frontière efficiente ex-post ainsi que les performances de ces trois stratégies sont représentées sur la figure 9.11.

La comparaison des performances de ces trois stratégies présentées dans le tableau 9.7 laisse un peu perplexé.

Nous pensions abaisser le niveau de risque, mais en fait c'est le modèle classique (covariance historique) qui fournit selon ce critère le meilleur résultat.

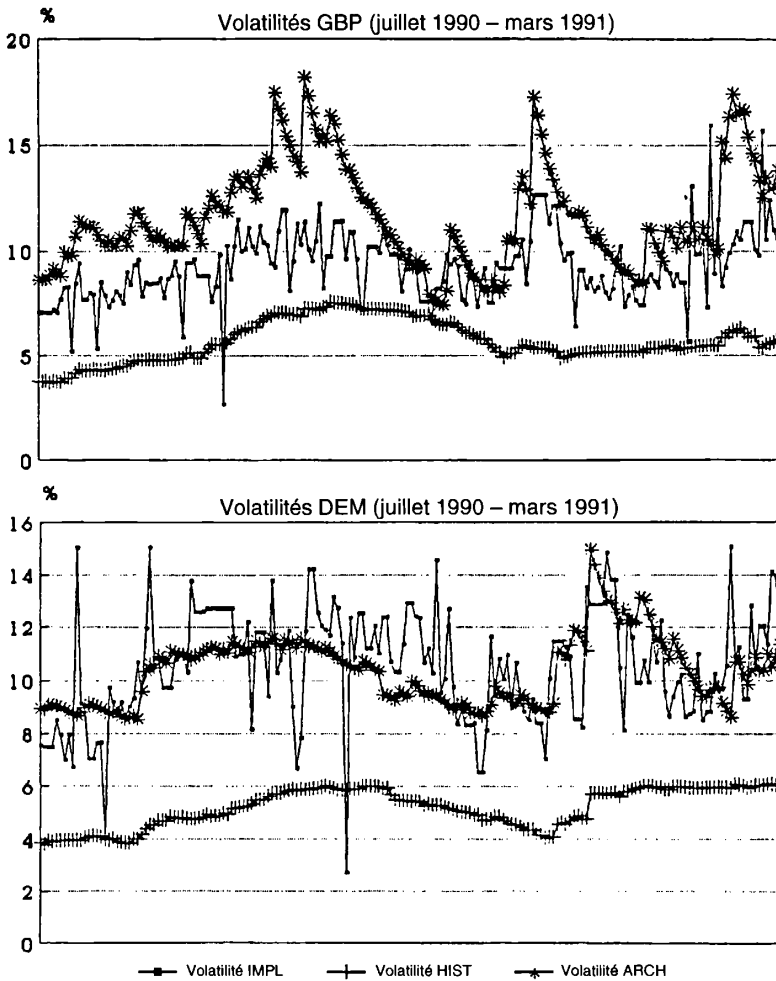


Figure 9.11: Volatilités GBP et DEM , juillet 1990 – mars 1991

Tableau 9.7: Comparaison des performances de trois stratégies

	Historique	GARCH	TARCH
Rentabilité	10.6%	13.0%	13.3%
Volatilité	17.8%	19.9%	19.6%
Ratio de Sharpe	0.033	0.151	0.168

En revanche les ratios de Sharpe des deux stratégies sont plus élevés que ceux de la stratégie classique. Il semble qu'il y ait bien de l'information exploitable

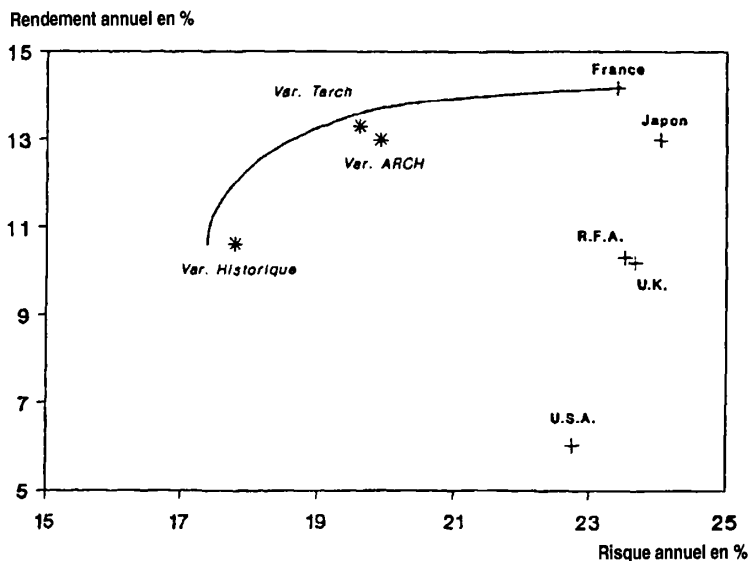


Figure 9.12

Investissement sur 5 marchés boursiers : frontière efficiente ex-post

dans les modèles (9.11) et (9.12), mais l'utilisation que nous en avons faite et/ou la mesure du risque que nous avons retenue ne mettent pas cette information à profit. Par ailleurs un coup d'œil aux structures de portefeuille (voir Boulier et Rabemananjara [1992]) montre qu'il faut souvent les modifier dans le contexte des stratégies ARCH, dont les performances avec coûts de transaction de 0.5% compris dominent celles de la stratégie classique.

La modélisation ARCH peut sans doute apporter des informations utiles pour le gestionnaire, notamment sur les mesures instantanées des expositions de son portefeuille (calcul de volatilités, bêta). Il est clair que l'approche des séries temporelles hétéroscédastiques peut être élargie à des modèles vectoriels plus ou moins élaborés - nous en avons vu un exemple. En revanche en ce qui concerne les performances, le risque et les stratégies de long terme, nos premiers résultats laissent un peu perplexe...

9.5 BILAN ET PERSPECTIVES

Les modèles ARCH sont venus « challenger » l'approche classique dans laquelle les actifs suivent des mouvements browniens géométriques. Ils sont en cela pour partie rivaux d'autres approches comme les lois stables de Lévy, les fractales et le mouvement brownien fractionnaire, le chaos déterministe et bien d'autres modélisations d'âge moins mûr. Les quelques quatre années de recherche et d'applications que nous venons d'illustrer ne suffisent pas à tirer des conclusions ni exhaustives ni définitives pour l'utilité pratique de cette approche.

Tout au plus viseront nous à faire un bilan provisoire en soulignant les indéniables apports, les limites et les difficultés.

Il est incontestable que les descriptions ARCH des variations de cours financiers sont plus réalistes et plus intuitives que l'approche classique. De plus un des principaux atouts par rapport aux modélisations alternatives réside dans la relative facilité d'utilisation des modèles ARCH. Ainsi le calcul du coût d'une option, la détermination de ratio de couverture (illustré dans Jamet et Rabemananjara par exemple) s'avèrent possibles, et dans le prolongement des techniques employées aujourd'hui. Bref, c'est le changement dans la bonne direction, sans être la révolution.

Toutefois les succès restent limités. C'est principalement dans le court terme (inférieur à 1 mois) que les effets ARCH sont sensibles et ils ne le sont vraiment que pour les actifs dérivés — prévision de volatilité implicite, calcul de volatilités effectives, arbitrages de court terme. Les résultats en gestion de portefeuille ou plus généralement sur les horizons de temps plus longs demeurent étranges. De plus l'usage de la modélisation se heurte à plusieurs difficultés : les estimations ne sont souvent pas stables dans le temps, les inefficiences ARCH ne sont pas très nombreuses et le phénomène reste la plupart du temps (hors période de krach par exemple) assez fin, si bien que la robustesse de cette modélisation est discutable... L'emploi avec succès de la « technologie » dérivée des modèles ARCH, nécessite donc beaucoup de précaution et une forte dose de conviction. Mais n'est-ce pas le lot de toute approche rationnelle des marchés financiers ?

Pour terminer sur une note positive et ne pas décourager les velléités d'applications, voici une liste de sujets pour lesquels les modèles ARCH pourraient à mon sens être adaptés. Tout d'abord les mesures de risque, notamment en gestion de portefeuille ou en terme de position : toutes les questions de replication et de gestion dynamique ; les positions de « noise trading » tirant avantage des sursréactions de marché ; l'analyse des mécanismes de contrôle de change, notamment dans le cadre du SME.

Nous ne saurions conclure sans souligner l'excellent travail de la recherche académique notamment Française qui a conçu cette modélisation ARCH, l'a développée ou inventé les outils d'estimation nécessaires. Après tout, si les succès en matière d'utilisation ne sont pas toujours flagrants, ce n'est peut-être pas dû au modèle de base...

BIBLIOGRAPHIE

- ANDEL, J., I. NETUKA et K. ZVARA [1984], On Threshold Autoregressive Processes, *Kybernetika*, **20**, 89-106.
- ASHLEY, R.A., D.M. PATTERSON et M.J. HINICH [1986], A Diagnostic Test for Nonlinear Serial Dependence in Time Series Fitting Errors, *Journal of Time Series Analysis*, **7**, 165-178.
- BABA, Y., R.F. ENGLE, D. KRAFT et K. KROMER [1989], *Multivariate Simultaneous Generalized ARCH*, Discussion Paper n° 89-57, University of San Diego.
- BACHELIER, L. [1900], Théorie de la spéculation, *Annales de l'Ecole Normale Supérieure*, 3e série, **17**, 21-86.
- BACHELIER, L. [1900], Theory of Speculation, reproduit in P. Cootner [1964], *The Random Character of Stock Market Prices*, pp.17-78, Cambridge (Mass.), The M.I.T. Press.
- BENGHABRIT, Y. et M. HALLIN [1992], Optimal Rank-based Tests Against First-Order Superdiagonal Bilinear Dependence, *Journal of Statistical Planning Inference*, **32**, 45-61.
- BERA, A. et S. LEE [1988], *Interaction Between Autocorrelation and Conditional Heteroskedasticity : A Random Coefficient Approach*, Discussion Paper, University of Illinois.
- BLACK, F. et M. SCHOLES [1973], The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, **81**, 637-659.
- BOLLERSLEV, T. [1986], Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307-327.
- BOLLERSLEV, T., R.Y. CHOU, N. JAYARAMAN et K.F. KRONER [1991], Les modèles ARCH en finance : un point sur la théorie et les résultats empiriques, *Annales d'Economie et de Statistique*, **24**, 1-59.
- BOLLERSLEV, T., R.Y. CHOU et K.F. KRONER [1992], ARCH Modeling in Finance : A Review of the Theory and Empirical Evidence, *Journal of Econometrics*, **52**, 5-59.
- BOLLERSLEV, T. et R.F. ENGLE [1989], *Common Persistence in Conditional Variances*, Mimeo, University of San Diego.
- BOLLERSLEV, T. [1987], *Modelling the Coherence in Short Run Nominal Exchange Rates : A Multivariate Generalized ARCH Model*, Mimeo, Northwestern University.
- BOLLERSLEV, T., R.F. ENGLE et J. WOOLDRIDGE [1988], A Capital Asset Pricing Model with Time-Varying Covariances, *Journal of Political Economy*, **96**, 116-131.
- BONOMO, M. et R. GARCÍA [1991], *Consumption and Equilibrium Asset Pricing : An Empirical Assessment*, Cahier 2991, CRDE, Université de Montréal.
- BORSCH-SUPPAN, A. et V.A. HAJIVASSILIOU [1990], *Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models*, Discussion Paper n° 960, Cowles Foundation.
- BOUGEROL, P. et N. PICARD [1992], Stationarity of GARCH Processes and of Some Non-Negative Time Series, *Journal of Econometrics*, **37**.
- BOULIER, J.-F. et R. RABEMANANJARA [1992], *Portfolio optimization and multivariate latent factor ARCH model*, FUR Conference, Cachan, Juin.
- BOX, G.E.P. et G.M. JENKINS [1976], *Time Series Analysis, Forecasting and Control*, San Francisco, Holden-Day, (édition révisée).

- BRAND, C. [1991], *Options de change à volatilités stochastiques*, Mémoire ISFA - CCF.
- CAMPBELL, J.Y. [1992], *Intertemporal Asset Pricing Without Consumption Data*, Working Paper 3989, NBER.
- CHAMBERLAIN, G. [1987], Asymptotic Efficiency in Estimation with Conditional Moment Restrictions, *Journal of Econometrics*, **34**, 305-334
- CHAN, K.S. et H. TONG [1986], On Estimating Thresholds in Autoregressive Models, *Journal of Times Series Analysis*, **7**, 179-190.
- CONNOR, G. et R.A. KORAJCZYK [1989], An Intertemporal Equilibrium Beta Pricing Model, *The Review of Financial Studies*, **2**, 373-392.
- DAVIDSON, R. et J.G. MACKINNON [1981], Several Tests for Model Specification in the Presence of Alternative Hypotheses, *Econometrica*, **49**, 781-793.
- DE GOOIJER, J.G., B. ABRAHAM, A. GOULD et L. ROBINSON [1985], Methods for Determining the Order of an Autoregressive-moving Average Process : A Survey, *International Statistical Review*, **53**, 301-329.
- DE GOOIJER, J.G. et K. KUMAR [1992], Some Recent Developments in Non-linear Time Series Modelling, Testing and Forecasting, *International Journal of Forecasting*, **8**.
- De HAAN, L., S.I. RESNICK, H. ROOTZEN et C. de VRIES [1989], Extremal Behaviour of Solutions to a Stochastic Difference Equation with Applications to ARCH Processes, *Stochastic Processes and their Applications*, **32**.
- DEMPSTER, A.P.N., N.M. LAIRD et D.B. RUBIN [1977], Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society, Serie B*, **39**, 1-38.
- DIEBOLD, F. et M. NERLOVE [1989], The Dynamics of Exchange Rate Volatility : A Multivariate Latent Factor ARCH Models, *Journal of Applied Econometrics*, **4**, 1-21.
- DROESBEKE, J.-J., B. FICHET et Ph. TASSI (Eds.) [1989], *Séries chronologiques : théorie et pratique des modèles ARIMA*, Paris, Economica.
- DUAN, J.-C. [1990], *The GARCH Option Pricing Model*, Preprint, Faculty of Management, McGill University.
- DUFFIE, D. et K.F. SINGLETON [1989], *Simulated Moments Estimation of Markov Models of Asset Prices*, Mimeo, Institution ? ? ? ?.
- DYBVIG, P. et S. ROSS [1985], The Analytics of Performance Measurement Using a Security Market Line, *The Journal of Finance*, **40**, 401-416.
- EATWELL, J., M. MILGATE et P. NEWMAN, (Eds.) [1989], *The New Palgrave : Finance*, London, Macmillan.
- EL BABSIRI, M. et J.-M. ZAKOÏAN [1990], *Approximation en temps continu d'un modèle ARCH à seuil*, Preprint, INSEE.
- ENGLE, R.F. [1982], Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, **50**, 987-1007.
- ENGLE, R.F. [1987], *Multivariate ARCH with Factors Returns-Cointegration in Variance*, Mimeo, University of San Diego.
- ENGLE, R.F. et T.P. BOLLERSLEV [1986], Modelling the Persistence of Conditional Variances, *Econometric Review*, **5**, 1-50.
- ENGLE, R.F. et D. KRAFT [1983], Multiperiod Forecast Error Variances of Inflation Estimated from ARCH Models, in A. Zellner (Ed.), *Applied Time Series Analysis of Economic Data*, Washington DC, Bureau of Census, 293-302.
- ENGLE, R.F., D. LILLIEN et R. ROBBINS [1987], Estimating Time Varying Risk Premia in the Term Structure : The ARCH-M Model, *Econometrica*, **55**, 391-407.

- ENGLE, R.F., V. NG et M. ROTHSCHILD [1989], A Factor ARCH Model for Stock Returns, Discussion Paper n° 89-31, University of San Diego.
- ENGLE, R.F., V. NG et M. ROTHSCHILD [1990], Asset Pricing with a Factor ARCH Covariance Structure : Empirical Estimators for Treasury Bills, *Journal of Econometrics*, **45**, 213-237.
- EPSTEIN, L. G. et S. ZIN [1989], Substitution, Risk Aversion, and the Temporal Behaviour of Consumption and Asset Returns : A Theoretical Framework, *Econometrica*, **57**, 937-969.
- ETHIER, S. et T. KURTZ [1986], *Markov Processes, Characterization and Convergence*, New York, John Wiley and Sons.
- EVANS, M. et P. WACHTEL [1990], *A Modern Look at Asset Pricing and Short-Term Interest Rates*, Working Paper 3245, NBER.
- EYSSARTIER, D. et N. LE DOUAREC [1989], *Application des séries temporelles à l'analyse financière*, Mémoire de stage, Ecole Polytechnique - CCF.
- FAMA, E.F. [1976], *Foundations of Finance*, New York, Basic Books.
- FÖLLMER, H. et M. SCHWEIZER [1993], *A Microeconomic Approach to Diffusion Models for Stock Prices*, Preprint, Universités de Bonn et Göttingen.
- GALLANT, R. et G. TAUCHEN [1989], Semi-nonparametric Estimation of Conditionally Constrained Heterogenous Processes : Asset Pricing Applications, *Econometrica*, **57**, 1091-1120.
- GARD, T. [1988], *Introduction to Stochastic Differential Equations*, New York, Marcel Dekker.
- GIBBONS, M. et W. FERSON [1985], Testing Asset Pricing Models With Changing Expectations and an Unobservable Market Portfolio, *Journal of Financial Economics*, **13**.
- GIOVANNINI, A. et P. JORION [1989], *Time-Series Tests of A Non-Expected-Utility Model of Asset Pricing*, Mimeo, Columbia University.
- GOURIEROUX, Ch. [1992], *Modèles ARCH : Applications financières et monétaires*, Paris, Economica.
- GOURIEROUX, Ch. et A. MONFORT [1989], *Statistique et modèles économétriques*, 2 volumes, Paris, Economica.
- GOURIEROUX, Ch. et A. MONFORT [1990], *Séries temporelles et modèles dynamiques*, Paris, Economica.
- GOURIEROUX, Ch. et A. MONFORT [1991a], Simulation Based Econometrics in Models with Heterogeneity, *Annales d'Economie et de Statistique*, n° 20/21.
- GOURIEROUX, Ch. et A. MONFORT [1991b], *Simulation Based Inference : a Survey with Special Reference to Panel Data Models*, Document de travail n° 9106, CREST/Département de la Recherche, INSEE (à paraître dans *The Journal of Econometrics*).
- GOURIEROUX, Ch. et A. MONFORT [1992a], Qualitative Threshold ARCH Models, *Journal of Econometrics*, **52**, 159-199.
- GOURIEROUX, Ch. et A. MONFORT [1992b], *Testing, Encompassing and Simulating Dynamic Econometric Models*, Document de travail n° 9214, CREST/Département de la Recherche, INSEE.
- GOURIEROUX, Ch., A. MONFORT et E. RENAULT [1991a], *Modèles dynamiques à facteurs*, CREST/Département de la Recherche, INSEE.
- GOURIEROUX, Ch., A. MONFORT et E. RENAULT [1991b], *Dynamic Factor Models*, Cahier n° 91.e, GREMAQ, Université de Toulouse.

- GOURIEROUX, Ch., A. MONFORT et E. RENAULT [1991c], Tests sur le rang, l'image et le noyau des coefficients de la matrice d'un système linéaire multivarié (à paraître dans *Annales d'Economie et de Statistique*).
- GOURIEROUX, Ch., A. MONFORT et E. RENAULT [1992], *Indirect Inference*, Document de travail n° 9215, CREST/Département de la Recherche, INSEE, (à paraître dans *The Journal of Applied Econometrics*)
- GOURIEROUX, Ch., A. MONFORT, E. RENAULT et A. TROGNON [1987], Simulated Residuals, *Journal of Econometrics*, **34**, 201-252.
- GOURIEROUX, Ch., A. MONFORT et A. TROGNON [1984], Pseudo Maximum Likelihood Methods : Theory, *Econometrica*, **52**, 681-700.
- GRANGER, C.W.J. et A. ANDERSEN [1978], *An Introduction to Bilinear Time Series Models*, Göttingen, Vanderhoeck and Ruprecht.
- GRANGER, C.W.J. et O. MORGENSTERN [1970], *Predictability of Stock Market Prices*, Lexington (Mass.), Lexington Books.
- GRANGER, C.W.J. et P. NEWBOLD [1976], Forecasting Transformed Series, *Journal of the Royal Statistical Society, Serie B*, **38**, 189-203.
- GRAR, A. [1992], *Hétéroscédasticité des taux de rentabilité des actifs financiers : une étude empirique sur le marché parisien*, Discussion Paper, CREST.
- GUEGAN, D. [1981], Etude d'un modèle non linéaire, le modèle superdiagonal d'ordre 1, *Comptes Rendus de l'Académie des Sciences de Paris*, **293**, Série I, 95-98.
- GUEGAN, D. [1987], Different Representations for Bilinear Models, *Journal of Time Series Analysis*, **8**, 389-408.
- GUEGAN, D. et T.D. PHAM [1992], Power of the Score Test Against Bilinear Time Series Models, *Statistica Sinica*, **2**, 157-169.
- GUIVARCH, Y. [1991], Sur une extension de la notion de loi semi-stable, *Annales de l'IHP*, **26**, 261-286.
- HAGGAN, V. et T. OZAKI [1981], Modelling Nonlinear Random Vibrations Using Amplitude-Dependent Autoregressive Time Series Models, *Biometrika*, **68**, 189-196.
- HAIJIVASSILIOU, V.A. et D. MCFADDEN [1989], *Country Heterogeneity and External Debt Crisis : Estimation by the Method of Simulated Moments*, Mimeo, Institution ???.
- HAMILTON, J. [1988], Rational Expectations Econometric Analysis of Changes in Regime : An Investigation of the Term Structure of Interest Rates, *Journal of Economic Dynamic and Control*, **12**, n° 213, 385-423.
- HAMILTON, J. [1989], A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycles, *Econometrica*, **57**, 357-384.
- HANSEN, L.P. [1982], Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, **50**, 1029-1054.
- HANSEN, L.P. et K.J. SINGLETON [1983], Stochastic Consumption, Risk Aversion and the Temporal Behaviour of Asset Returns, *Journal of Political Economy*, **91**, 249-265.
- HARVEY, A., E. RUIZ et N. SHEPHARD [1992], *Multivariate Stochastic Variance Models*, Discussion Paper n° 132, London School of Economics.
- HINICH, M.J. [1982], Testing for Gaussianity and Linearity of a Stationary Time Series, *Journal of Time Series Analysis*, **3**, 169-176.
- HUBERMAN, G., S. KANDEL et R. STAMBAUGH [1987], Mimicking Portfolios and Exact Arbitrage Pricing, *The Journal of Finance*, **42**, 1-9.
- HULL J. et A. WHITE [1987], The Pricing of Options on Assets with Stochastic Volatilities, *The Journal of Finance*, **42**, n° 2, 281-300.

- INGRAM, B. et B. LEE [1988], Estimation by Simulation, Mimeo, Cornell University (à paraître dans *The Journal of Econometrics*).
- JAMET, F. et R. RABEMANANJARA [1989], *Prévision de volatilité et modèles ARCH*, Journées AFFI, Paris, Décembre.
- JOBSON, J. et B. KORKIE [1982], Potential Performance and Tests of Portfolio Efficiency, *Journal of Financial Economics*, **10**, 433-466.
- Journal of Finance. Vol XL II, n° 2.
- KEANE, M.P. [1990], *A Computationally Practical Simulation Estimator for Panel Data, with Applications to Estimating Temporal Dependence in Employment and Wages*, Discussion Paper, University of Minnesota.
- KEENAN, D. M. [1985], A Tukey Nonadditivity-type Test for Time Series Nonlinearity, *Biometrika*, **72**, 39-44.
- KESTEN, H. [1973], Random Difference Equations and Renewal Theory for Products of Random Matrices, *Acta Mathematica*, **131**, 208-248.
- KIM, K. et P. SCHMIDT [1989], *Unit Root Tests with Conditional Heteroskedasticity*, Discussion Paper, Michigan State University.
- KING, M., E. SENTANA et S. WADHWANI [1990], *A Heteroscedastic Factor Model of Asset Returns and Risk Premia with Time-Varying Volatility: An Application to Sixteen World Stock Markets*, London School of Economics.
- KITAGAWA, G. [1987], Non-Gaussian State-Space Modeling of Nonstationary Time Series, *Journal of the American Statistical Association*, **82**, 1032-1063
- KUSHNER, H. [1984], *Approximation and Weak Convergence Methods for Random Processes*, Cambridge (Mass.), The M.I.T. Press.
- LAFFONT, J.J. [1983], *Fondements de l'économie publique*, vol.2: *Economie de l'Incertain et de l'Information*, Paris, Economica.
- LAMBERTON, D. et B. LAPEYRE [1991], *Introduction au calcul stochastique appliqué à la finance, Mathématiques et Applications*, Paris, Editions Ellipses.
- LAROQUE, G. et B. SALANIÉ [1989], Estimation of Multi-Market Fix Price Models: An Application of Pseudo Maximum Likelihood Methods, *Econometrica*, **57**, 831-860.
- LAROQUE, G. et B. SALANIÉ [1990], *Estimating the Canonical Disequilibrium Model: Asymptotic Theory and Finite Sample Properties*, Document de Travail n° 9005, CREST-Département de la Recherche, INSEE.
- LAROQUE, G. et B. SALANIÉ [1992], *Simulation Based Estimation of Models with Lagged Latent Variables*, Mimeo, Département de la Recherche, INSEE.
- LAWRANCE, A. J. et P.A.W. LEWIS [1985], Modelling And Residual Analysis of Nonlinear Autoregressive Time Series in Exponential Variables, *Journal of the Royal Statistical Society, Serie B*, **47**, 165-183 [with discussion 183-202].
- LERMAN, S. et C. MANSKI [1981], On the Use of Simulated Frequencies to Approximate Choice Probabilities, in C. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge (Mass.), The M.I.T. Press, 305- 319.
- LIN, W.L. [1991], *Alternative Estimators for Factor GARCH Models: a Monte Carlo Comparison*, University of Wisconsin.
- LINTNER, J. [1965], The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets, *Review of Economics and Statistics*, **47**, 13-37.
- LIU, J. [1989], A Simple Condition for the Existence of Some Stationary Bilinear Time Series, *Journal of Time Series Analysis*, **10**, 33-39.

- LIU, J. [1990], A Note on Causality and Invertibility of a General Bilinear Time Series Model, *Advanced Applied Probability*, **22**, 247-250.
- LIU, J. et P.J. BROCKWELL [1988], On the general bilinear time series model, *Journal of Applied Probability*, **25**, 553-564.
- LUCAS, R.E. [1978], Asset Prices in an Exchange Economy, *Econometrica*, **46**, 1426-1445.
- MARKOWITZ, H.M. [1952], Portfolio Selection, *Journal of Finance*, **7**, 77-91.
- MARKOWITZ, H.M. [1959], *Portfolio Selection : Efficient Diversification of Investments*, Yale University Press, New Haven. Reprinted : New York, John Wiley and Sons, 1970.
- MCFADDEN, D. [1989], A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration, *Econometrica*, **57**, 995-1026.
- MCLEOD, A.I. et W.K. LI [1983], Diagnostic Checking ARMA Time Series Models Using Squared-Residual Autocorrelations, *Journal of Time Series Analysis*, **4**, 269-273.
- MELARD, G. [1984], Algorithm AS197 : A Fast Algorithm for the Exact Likelihood of Autoregressive-Moving Average Models, *Journal of the Royal Statistical Society, Serie C Applied Statistics*, **33**, 104-114.
- MELARD, G. [1985], *Analyse de données chronologiques*, Collection Séminaire de mathématiques supérieures - Séminaire scientifique OTAN (NATO Advanced Study Institute) n° 89, Montréal, Presses de l'Université de Montréal.
- MELARD, G. [1990], *Méthodes de prévision à court terme*, Coll. Statistique et mathématiques appliquées, Bruxelles, Editions de l'Université de Bruxelles, et Paris, Editions Ellipses.
- MELARD, G. et R. ROY [1988], Modèles de séries chronologiques avec seuil, *Revue de Statistique Appliquée*, **36**, 5-24.
- MILHOJ, A. [1985], The Moment Structure of ARCH Processes, *Scandinavian Journal of Statistics*, **12**, 281-292.
- MODIGLIANI, F. et M.H. MILLER [1958], The Cost of Capital, Corporation Finance and the Theory of Investment, *American Economic Review*, **48**, 261-297.
- MOSSIN, J. [1966], Equilibrium in a Capital Asset Market, *Econometrica*, **34**, 768-783.
- NELSON, D.B. [1990a], ARCH Models as Diffusion Approximations, *Journal of Econometrics*, **45**, 7-38.
- NELSON, D.B. [1990b], Stationarity and Persistence in the GARCH(1,1) Model, *Econometric Theory*, **6**, 318-334.
- NEVEU, J. et E. PARDOUX [1990], *Modèles de diffusion*, Cours de l'Ecole Polytechnique.
- NEWBY, W. [1989], *Efficient Instrumental Variables Estimation of Nonlinear Models*, Princeton, Princeton University Press.
- NEWBY, W. et K. WEST [1987], A Simple Positive, Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, **55**, 703-708.
- PAGAN, A. et G. SCHWERT [1989], *Alternative Models for conditional Stock Volatility*, Working Paper 2955, NBER.
- PAKES, A. et D. POLLARD [1989], Simulation and the Asymptotics of Optimization Estimators, *Econometrica*, **57**, 1027-1057.
- PETRUCCELLI, J.D. et N. DAVIES [1986], A Portmanteau Test for Self-exciting Threshold-type Nonlinearity in Time Series, *Biometrika*, **73**, 687-694.
- PASTORELLO, S. et E. RENAULT [1992], *La notion de Facteur en Finance*, Document de travail, GREMAQ, Université des Sciences Sociales, Toulouse.

- PETRUCCCELLI, J. D. et S.W. WOOLFORD [1984], A Threshold AR(1) Model, *Journal of Applied Probability*, **21**, 270-286.
- PLUCHET, V et A. SAUVANT [1990], *Modèles hétéroscédastiques et pricing d'options*, Mémoire de stage, Ecole Polytechnique - CCF.
- PRIESTLEY, M.B. [1980], State-dependent Models : A General Approach to Non-linear Time Series Analysis, *Journal of Time Series Analysis*, **1**, 47-71.
- PRIESTLEY, M.B. [1988], *Non-Linear and Non-Stationary Time Series Analysis*, New York, Academic Press.
- RABEMANANJARA, R. et J.-M. ZAKOÏAN [1991a], *Arch Models and Asymmetries in Volatility*, Conference on non linear dynamics and econometrics, Los Angeles, April.
- RABEMANANJARA, R. et J.-M. ZAKOÏAN [1991b], *Threshold GARCH Models and Asymmetries in Volatility*, Preprint, INSEE.
- RABEMANANJARA, R. et N. BATT [1991], *Distribution empirique des rendements des actions et modèles ARCH*, Congrès international de l'AFFI, Louvain.
- RENAULT, E. et N. TOUZI [1992], *Option Hedging and Implicit Volatilities*, Document de travail, CREST-ENSAE.
- SCOTT, L.O. [1987], Option pricing when the variance changes randomly, *Journal of Financial and Quantitative analysis*, **22**.
- SHARPE, W.F. [1964], Capital Asset Prices : A Theory of Market equilibrium under Conditions of Risk, *The Journal of Finance*, **19**, 425-442.
- SOLOMON, E. [1967], *The Theory of Financial Management*, New York, Columbia University Press.
- STEIN E.M. et J.C. STEIN [1991], Stock Price Distributions with Stochastic Volatility : An Analytic Approach, *The Review of Financial Studies*, **4**.
- STROOCK, D.W. et S.R.S. VARADHAN [1979], *Multidimensional Diffusion Processes*, New York, Springer-Verlag.
- SUBBA RAO, T. [1981], On the theory of bilinear models, *Journal of the Royal Statistical Society, Serie B*, **42**, 245-292.
- SUBBA RAO, T. et M.M. GABR [1980], A test for Linearity of Stationary Time Series, *Journal of Time Series Analysis*, **1**, 145-158.
- TAUCHEN, G. et R. HUSSEY [1991], Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models, *Econometrica*, **59**, 371-396.
- TAYLOR, S. [1986], *Modelling Financial Time Series*, New York, John Wiley and Sons.
- TIAO, G.C. et G.E.P. BOX [1981], Modeling Multiple Time Series with Applications, *Journal of American Statistical Association*, **76**, 802-816.
- TOBIN, J. [1958], Liquidity Preference as Behavior Towards Risk, *Review of Economic Studies*, **25**, 65-86.
- TONG, H. [1978], On a Threshold Model, in C.H. Chen (Ed.), *Pattern Recognition and Signal Processing*, Alphen aan den Rijn, Sijthoff and Norrdhoff, 575-586.
- TONG, H. [1983], *Threshold Models in Non-linear Time Series Analysis*, New York, Springer-Verlag.
- TONG, H. [1990], *Non-linear Time Series : A Dynamical System Approach*, Oxford, Oxford University Press.
- TSAY, R.S. [1991], Detecting and Modeling Nonlinearity in Univariate Time Series Analysis, *Statistica Sinica*, **1**, 431-451.
- TSAY, R.S. et G.C. TIAO [1984], Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA

- Models, *Journal of American Statistical Association*, **79**, 84-96.
- VAN PRAAG, B.M.S. et J.P. HOP [1987], *Estimation of Continuous Models on the Basis of Set-Valued Observations*, papier présenté au Congrès de la Société Européenne d'Econométrie.
- WEISS, A. [1984], ARMA Models with ARCH Errors, *Journal of Time Series Analysis*, **5**, 129-143.
- WEISS, A. [1986], ARCH et Bilinear Time Series Models : Comparaison and Combination, *Journal of Business and Economic Statistics*, **4**, 59-70.
- WHITE, H. [1980], A Heteroskedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, **48**, 817-838.
- WHITE, H. [1982], Maximum Likelihood Estimation of Misspecified Models, *Econometrica*, **50**, 483-499.
- WIGGINS J.B. Option Values Under Stochastic Volatility, *Journal of Financial Economics*, **19**, 351-372.
- WOOLDRIDGE, J. [1991], On the Application of Robust Regression Based Diagnostics to Models of Conditional means and Conditional Variances, *Journal of Econometrics*, **47**, 5-46.
- ZAKOÏAN, J.-M. [1991], *Models ARCH à seuils*, Thèse, Université Paris IX.
- ZAKOÏAN, J.-M. [1992], Modèles ARCH : une revue de la littérature, *Journal de la Société Statistique de Paris*, **133**, 40-57.

Liste des Tableaux

2.1	Autocorrélations et autocorrélations partielles de la série DPIB . . .	32
2.2	Autocorrélations et autocorrélations partielles de la série XPIB . . .	33
2.3	Estimation d'un modèle simple sur la série XPIB	33
2.4	Autocorrélations et autocorrélations partielles de la série DTAUR .	35
2.5	Estimation d'un premier modèle sur la série DTAUR	35
2.6	Autocorrélations des résidus du premier modèle	35
2.7	Estimation d'un second modèle sur la série DTAUR	35
2.8	Autocorrélations des résidus du second modèle	36
2.9	Estimation d'un troisième modèle sur la série DTAUR	36
2.10	Autocorrélations des résidus du troisième modèle	36
3.1	Distribution empirique des rentabilités	60
3.2	Modèles estimés	61
4.1	Test de bruit blanc	90
6.1	Mouvement brownien	128
6.2	Processus d'Ornstein Uhlenbeck	130
9.1	Estimation des paramètres 1976-1990	210
9.2	Estimation des paramètres du modèle GARCH(1,1)	211
9.3	Coefficients de corrélation de la volatilité implicite	212
9.4	Performances des stratégies d'investissement fondées sur trois modèles	214
9.5	Valeur du Call suivant 4 méthodes de calcul et 3 hypothèses de leur prix d'exercice K	217
9.6	Test des stratégies d'arbitrage sur le marché des options du CAC 40	219
9.7	Comparaison des performances de trois stratégies	222

3.3	Évolution de la volatilité : Indice pondéré avec dividende (GARCH(1,1))	63
3.4	Évolution de la volatilité : Indice pondéré sans dividende (GARCH(1,1))	63
3.5	Régions de persistance et de stationnarité	66
4.1	Intervalles de prévision pour l'estimation ARMA	84
4.2	Intervalles de prévision pour l'estimation ARMA-GARCH	85
4.3	Stratégies d'intervention fondées sur un modèle ARMA et sur un modèle ARMA-GARCH	86
4.4	Volatilité estimée	95
4.5	Volatilités estimées $\hat{\beta}_{ii}$	96
4.6	Coefficients estimés $\hat{\beta}_{ij}$	97
4.7	Influence de Y_{t-1} sur la volatilité	98
6.1	Brownien géométrique : densité des estimateurs de μ	129
6.2	Brownien géométrique : densité des estimateurs de σ	129
6.3	Processus d'Ornstein Uhlenbeck : densité des estimateurs de k	131
6.4	Processus d'Ornstein Uhlenbeck : densité des estimateurs de a	131
6.5	Processus d'Ornstein Uhlenbeck : densité des estimateurs de σ	132
9.1	Distribution des skewness des valeurs	205
9.2	Distribution des kurtosis des valeurs	205
9.3	Distribution d'une loi normale et distribution empirique de l'indice CAC pour des données quotidiennes, hebdomadaires et mensuelles (1976–1990)	206
9.4	Rendements quotidiens de l'indice CAC Général en 1987 et volatilité estimée sous l'hypothèse du modèle GARCH(1,1)	207
9.5	Rendements quotidiens de l'indice CAC de 1976 à 1990 : distributions de e_t/σ_t et de e_t/σ	208
9.6	Fonctions d'autocorrélation : séries de résidus	209
9.7	Volatilités	212
9.8	Comparaison en prévision des modèles : prévision de l'indice CAC	214
9.9	Volatilité effective	218
9.10	Kurtosis d'un portefeuille en fonction du nombre de valeurs : données hebdomadaires 1981–1989 ; tri par capitalisation boursière	220
9.11	Volatilités GPB et DEM , juillet 1990 – mars 1991	222
9.12	Investissement sur 5 marchés boursiers : frontière efficiente ex-post	223

TABLE DES MATIÈRES

AVANT-PROPOS	9
CHAPITRE 1	
LES MODÈLES ARCH ET LEUR HISTOIRE (Jean-Jacques DROESBEKE, Bernard FICHET, Philippe TASSI).....	11
CHAPITRE 2	
MODÈLES LINÉAIRES ET NON LINÉAIRES (Guy MÉLARD).....	17
2.1 INTRODUCTION	17
2.2 SÉRIES TEMPORELLES ET PROCESSUS STOCHASTIQUES	18
2.2.1 Processus stochastiques	18
2.2.2 Processus stationnaires.....	19
2.2.3 Processus bruit blanc	20
2.2.4 Classes de processus	21
2.2.5 Etude des processus ARMA	24
2.3 MODÈLES LINÉAIRES	27
2.3.1 Description de la méthode de Box et Jenkins.....	27
2.3.2 Exemple 1	31
2.3.3 Exemple 2	34
2.4 MODÈLES NON LINÉAIRES	37
2.4.1 Modèles bilinéaires	37
2.4.2 Modèles autorégressifs à seuils	41
2.4.3 Modèles autorégressifs conditionnellement hétéroscédastiques ..	43
2.4.4 Quelques autres modèles non linéaires	48
2.5 TESTS DE LINÉARITÉ.....	50
2.5.1 Tests basés sur les autocorrélations	50
2.5.2 Tests basés sur la régression.....	51
2.5.3 Autres tests	52
CHAPITRE 3	
LES MODÈLES HÉTÉROSCÉDASTIQUES UNIVARIÉS (Christian GOURIEROUX).....	53
3.1 DIVERSES MODÉLISATIONS	53
3.1.1 Modèles ARCH(q) (Engle [1982]).....	53
3.1.2 Modèles GARCH (p, q) (Bollerslev [1986])	54

3.1.3	Modèles ARMAX-GARCH.....	55
3.1.4	Modèles ARCH-M (Engle-Lillien-Robbins [1987])	55
3.1.5	Modèles autorégressifs à coefficients aléatoires (Bera-Lee [1989])	55
3.1.6	Formulations non paramétriques	56
3.2	STATIONNARITÉ	57
3.2.1	Stationnarité faible	57
3.2.2	Valeurs rencontrées en pratique	59
3.2.3	Stationnarité forte	61
3.3	QUEUES DE DISTRIBUTION.....	67
3.3.1	Lien entre kurtosis conditionnelle et kurtosis marginale	67
3.3.2	Kurtosis et choix de l'unité de temps.....	68
3.3.3	Normalité conditionnelle et choix de l'unité de temps.....	70
CHAPITRE 4		
ESTIMATION, PRÉVISION ET TESTS : Modèles univariés		
	(Christian GOURIEROUX).....	71
4.1	ESTIMATION LORSQUE LES MOMENTS SONT PARAMÉTRÉS	72
4.1.1	Méthodes du pseudo-maximum de vraisemblance	72
4.1.2	Exemples	75
	<i>i) Modèles conditionnellement normaux</i>	75
	<i>ii) Paramétrages indépendants de la moyenne et de la variance</i>	75
	<i>iii) Modèle de régression avec erreurs ARCH gaussiennes ...</i>	76
4.1.3	Méthode en deux étapes	77
4.1.4	Comparaison des deux approches	78
	<i>i) Les modèles ARCH-M.....</i>	78
	<i>ii) Efficacité relative</i>	79
4.1.5	Méthode des moments généralisés	80
4.2	PREVISION.....	82
4.2.1	Forme des intervalles de prévision.....	82
4.2.2	Un exemple.....	83
4.2.3	Volatilité et stratégies d'intervention.....	85
4.3	TESTS D'HYPOTHÈSES	86
4.3.1	Hypothèse de marche aléatoire	87
	<i>i) Importance de l'hypothèse en finance</i>	87
	<i>ii) La procédure de test classique</i>	89
	<i>iii) Test portmanteau corrigé.....</i>	90
4.3.2	Test d'homoscédasticité.....	92
4.4	DU GENERAL AU SPECIFIQUE	94
4.4.1	Le principe	94

4.4.2 Un exemple (voir Gouriéroux et Monfort [1992]).....	94
<i>i) Asymétrie de la volatilité (« leverage effect »).....</i>	95
<i>ii) Recherche d'une spécification contrainte.....</i>	96
<i>iii) Estimation de la spécification partiellement contrainte ...</i>	98
<i>iv) Estimation d'une formulation GARCH(1,1).....</i>	98

CHAPITRE 5

GÉNÉRALISATIONS ET ALTERNATIVES DE LA MODÉLISATION ARCH

(Alain MONFORT).....	99
5.1 INTRODUCTION.....	99
5.2 PROCESSUS ARCH ET GARCH MULTIVARIÉS.....	99
5.2.1 Définitions générales.....	99
5.2.2 Modèles dynamiques à facteurs.....	101
5.2.3 Modèles GARCH à facteurs endogènes.....	103
5.2.4 Modèles GARCH à facteurs exogènes.....	104
5.2.5 Estimation des modèles GARCH à facteurs.....	105
5.3 MODÈLES À FACTEURS DISCRETS ET MODÈLES À VOLATILITÉ STOCHASTIQUE DISCRÈTE.....	106
5.3.1 Définition.....	106
5.3.2 Filtrage, lissage, estimation.....	108
<i>i) Algorithme de type Kalman.....</i>	108
<i>ii) Algorithme EM.....</i>	109
5.4 MODÈLES À VOLATILITÉ STOCHASTIQUE CONTINUE.....	111
5.4.1 Cas univarié.....	111
5.4.2 Cas multivarié.....	113

CHAPITRE 6

INFÉRENCE FONDÉE SUR DES SIMULATIONS

(Alain MONFORT).....	115
6.1 INTRODUCTION.....	115
6.2 LA MÉTHODE DES MOMENTS SIMULÉS.....	116
6.3 LA MÉTHODE DU MAXIMUM DE VRAISEMBLANCE SIMULÉE.....	118
6.4 LA MÉTHODE DU PSEUDO-MAXIMUM DE VRAISEMBLANCE SIMULÉE.....	119
6.5 MÉTHODES D'INFÉRENCE INDIRECTE.....	122
6.5.1 Présentation des méthodes.....	122
6.5.2 Remarques.....	123
6.5.3 Estimation des paramètres d'une équation différentielle stochastique (EDS).....	125

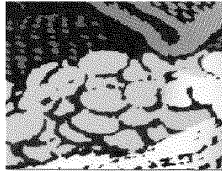
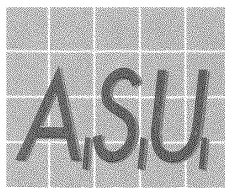
6.5.4 Estimation des paramètres d'un modèle à volatilité stochastique en temps continu.....	126
6.5.5 Applications numériques.....	127
i) <i>Cas du mouvement brownien géométrique</i>	128
ii) <i>Cas du processus d'Ornstein-Uhlenbeck</i>	130
CHAPITRE 7	
MODÈLES A FACTEURS EN FINANCE	
(Sergio PASTORELLO et Eric RENAULT).....	133
7.1 INTRODUCTION.....	133
7.2 LA NOTION DE FACTEUR EN FINANCE	134
7.2.1 Le problème de l'actualisation de revenus risqués	134
i) <i>L'aversion par rapport au risque</i>	134
ii) <i>Actifs et portefeuilles</i>	135
iii) <i>Choix de portefeuille optimal sur une période</i>	135
iv) <i>Aspect intertemporel des programmes</i> <i>de consommation-investissement</i>	136
v) <i>Un modèle d'évaluation synthétique : le modèle</i> <i>d'Epstein et Zin</i>	136
7.2.2 Les modèles à facteurs associés au modèle d'Epstein et Zin . . .	138
i) <i>La notion de relation multibêta</i>	138
ii) <i>Relation d'Epstein et Zin sous l'hypothèse</i> <i>de lognormalité</i>	139
iii) <i>Interprétations alternatives des facteurs</i>	140
iv) <i>Modèles à facteurs ne faisant pas intervenir</i> <i>la consommation</i>	141
7.3 CLASSIFICATION DES MODÈLES A FACTEURS	145
7.3.1 Le rôle des facteurs exogènes.....	146
i) <i>Modèles dynamiques à facteurs</i>	146
ii) <i>Endogénéisation statistique des facteurs exogènes</i>	149
7.3.2 Le rôle des facteurs endogènes	150
i) <i>Elimination de facteurs non observables</i>	150
ii) <i>La réduction de la dimension de l'incertitude</i>	154
7.4 APPLICATION A L'EVALUATION DE PERFORMANCE DE PORTEFEUILLES.....	158
7.4.1 Indice de performance de Sharpe	158
7.4.2 Performances conditionnelles et historiques.....	159
7.4.3 Tests d'efficience et de performance.....	161
Annexe A : Preuve du Théorème 7.1	164
Annexe B : Preuve du Théorème 7.4.....	169

Liste des Figures

2.1	Graphique de la série de longueur 400 générée par un processus bruit blanc et les autocorrélations de la série des 100 premières données .	25
2.2	Graphique de la série de longueur 400 générée par un processus AR(1) et les autocorrélations de la série des 100 premières données	25
2.3	Graphique de la série de longueur 400 générée par un processus MA(1) et les autocorrélations de la série des 100 premières données	26
2.4	Graphique de la série de longueur 400 générée par un processus ARMA(1,1) et les autocorrélations de la série des 100 premières données	26
2.5	Série artificielle avec variations de niveau	26
2.6	Série artificielle qui demande à être transformée en logarithmes . . .	27
2.7	Les données du produit intérieur brut	32
2.8	Les données de $\nabla \nabla_4$ PIB	32
2.9	Les données de PIB et de ses prévisions FPIB	34
2.10	Les séries TAUR et DTAUR	34
2.11	La série des résidus du troisième modèle	37
2.12	Série artificielle générée par un processus bilinéaire	40
2.13	Autocorrélations de la série artificielle de longueur 100 générée par un processus bilinéaire et des carrés de cette série	40
2.14	La série artificielle générée par un processus bilinéaire en fonction de la série retardée d'une unité	40
2.15	Série artificielle générée par un processus autorégressif à seuil	42
2.16	Autocorrélations de la série artificielle de longueur 100 générée par un processus autorégressif à seuil (à gauche) et la série en fonction de la série retardée d'une unité (à droite)	43
2.17	Série artificielle générée par un processus ARCH(1)	46
2.18	Autocorrélations de la série artificielle de longueur 100 générée par un processus ARCH(1) (à gauche) et de la série des carrés (à droite)	47
2.19	Série artificielle générée par un processus AR(1) à erreurs ARCH(1)	48
2.20	Autocorrélations de la série artificielle de longueur 100 générée par un processus AR(1) à erreurs ARCH(1)	49
3.1	Évolution de la volatilité : Indice équi pondéré avec dividende (GARCH(1,1))	62
3.2	Évolution de la volatilité : Indice équi pondéré sans dividende (GARCH(1,1))	62

CHAPITRE 8	
LES PROCESSUS ARCH COMME APPROXIMATIONS	
DE PROCESSUS EN TEMPS CONTINU	
(Laure ELIE).....	171
8.1 INTRODUCTION	171
8.2 CALCUL STOCHASTIQUE.....	173
8.2.1 Mouvement Brownien	173
8.2.2 Intégrale stochastique	174
8.2.3 Processus d'Itô.....	176
8.3 EQUATIONS DIFFERENTIELLES STOCHASTIQUES ET DIFFU SIONS	178
8.3.1 Définition et propriétés	178
8.3.2 Exemples de diffusions	180
8.3.3 Probabilité invariante	182
8.3.4 Le cas vectoriel	184
8.4 CONVERGENCE DE CHAINES DE MARKOV VERS DES DIFFU SIONS	185
8.4.1 Discrétisation d'Euler d'une diffusion	185
8.4.2 Théorème de convergence.....	185
8.5 LES PROCESSUS HÉTÉROSCEDASTIQUES COMME APPROXI MATIONS DE PROCESSUS EN TEMPS CONTINU	187
8.5.1 Modèle GARCH-M (D.B. Nelson [1990a]).....	188
8.5.2 Modèle ARCH exponentiel autorégressif d'ordre 1	194
8.5.3 Modèle GARCH à seuil (M. El Babsiri et J.M. Zakoian [1990])	195
8.6 CONVERGENCE DES LOIS STATIONNAIRES	196
8.7 CONCLUSION.....	200
CHAPITRE 9	
DE L'UTILITÉ DES MODÈLES ARCH	
(Jean-François BOULIER)	203
9.1 POUVOIR DESCRIPTIF	204
9.1.1 Modélisation GARCH de l'indice CAC.....	207
9.1.2 Modélisation GTARCH de l'indice CAC	208
9.1.3 Modélisation du MATIF	210
9.2 PREVOIR AVEC LES ARCH.....	211
9.2.1 Prévoir la volatilité du MATIF.....	211
9.2.2 Prévoir la volatilité des cours de change	212
9.2.3 Modèles GARCH-M pour prévoir le CAC.....	213
9.3 LES OPTIONS DANS UN UNIVERS HÉTÉROSCÉDASTIQUE . .	215
9.3.1 Modèle stochastique et valorisation d'option	215
9.3.2 Évaluation numérique	216

9.3.3 Volatilités effectives	217
9.3.4 Stratégie d'arbitrage des options CAC	218
9.4 GESTION DE PORTEFEUILLE ET MODÉLISATION ARCH	220
9.4.1 Persistence d'effets ARCH	220
9.4.2 Modèles à facteurs latents hétéroscédastiques	220
9.4.3 Simulation de portefeuille et performances	221
9.5 BILAN ET PERSPECTIVES	223
BIBLIOGRAPHIE	225
Liste de Tableaux	233
Liste de Figures	235
Table des Matières	237



Créée en 1971 pour promouvoir l'enseignement de la statistique et représenter ceux qui l'enseignent, l'Association des Statisticiens Universitaires (ASU) a rapidement dépassé son origine strictement universitaire, et a désormais vocation à accueillir tous les statisticiens d'expression française, comme le prouve sa nouvelle appellation depuis mai 1987: Association pour la Statistique et ses Utilisations. Forte de plus de quatre cents adhérents, l'ASU organise chaque année les Journées de Statistique, principal colloque en ce domaine, est chargée par la Société Statistique de France d'éditer la "Revue de Statistique Appliquée", et a pris l'initiative, en 1984, des Journées d'Etude en Statistique qui sont à l'origine de cette collection. En outre, les membres de l'ASU se réunissent au sein de groupes spécialisés organisés autour d'un thème ou d'un domaine d'application.

Cet ouvrage fait suite aux Journées d'Etude en Statistique organisées par l'ASU en octobre 1992 sur le thème des modèles ARCH et de leurs applications à la finance, au Centre International de Rencontres Mathématiques (CIRM) de Marseille-Luminy, avec le concours de la Société Mathématique de France (SMF).

Le but de ces Journées est de permettre à un public peu spécialisé de se consacrer à l'approfondissement d'un sujet précis, avec quatre orientations essentielles: l'acquisition des résultats de base, la présentation des développements les plus importants et les plus récents, les perspectives futures et les problèmes liés à l'application des théories et méthodes énoncées.

Dans la lignée de l'utilisation des processus stochastiques de type ARMA pour le traitement des séries chronologiques, la classe des modèles non linéaires ARCH (autorégressifs conditionnellement hétéroscédastiques) permet une étude complète, illustrée par des applications dans des domaines concernant principalement des problèmes financiers et monétaires.

ISBN 2-8004-1081-7



Règles d'utilisation de copies numériques d'œuvres littéraires publiées par les Editions de l'Université de Bruxelles et mises à disposition par les Bibliothèques de l'ULB

L'usage des copies numériques d'œuvres littéraires, ci-après dénommées « copies numériques », publiées par les Editions de l'Université de Bruxelles, ci-après dénommées EUB, et mises à disposition par les Bibliothèques de l'ULB, implique un certain nombre de règles de bonne conduite, précisées ici. Celui-ci est reproduit sur la dernière page de chaque copie numérique publiée par les EUB et mise en ligne par les Bibliothèques; il s'articule selon les trois axes [protection](#), [utilisation](#) et [reproduction](#).

Protection

1. Droits d'auteur

La première page de chaque copie numérique indique les droits d'auteur d'application sur l'œuvre littéraire.

La mise à disposition par les Bibliothèques de l'ULB de la copie numérique a fait l'objet d'un accord avec les EUB, notamment concernant les règles d'utilisation précisées ici.

Pour les œuvres soumises à la législation belge en matière de droit d'auteur, les EUB auront pris le soin de conclure un accord avec leurs ayants droits afin de permettre la mise en ligne des copies numériques.

2. Responsabilité

Malgré les efforts consentis pour garantir les meilleures qualité et accessibilité des copies numériques, certaines déficiences peuvent y subsister – telles, mais non limitées à, des incomplétudes, des erreurs dans les fichiers, un défaut empêchant l'accès au document, etc. -.

Les EUB et les Bibliothèques de l'ULB déclinent toute responsabilité concernant les dommages, coûts et dépenses, y compris des honoraires légaux, entraînés par l'accès et/ou l'utilisation des copies numériques. De plus, les EUB et les Bibliothèques de l'ULB ne pourront être mis en cause dans l'exploitation subséquente des copies numériques ; et la dénomination des EUB et des 'Bibliothèques de l'ULB', ne pourra être ni utilisée, ni ternie, au prétexte d'utiliser des copies numériques mises à disposition par eux.

3. Localisation

Chaque copie numérique dispose d'un URL (uniform resource locator) stable de la forme <http://digistore.bib.ulb.ac.be/annee/nom_du_fichier.pdf> qui permet d'accéder au document ; l'adresse physique ou logique des fichiers étant elle sujette à modifications sans préavis. Les bibliothèques de l'ULB encouragent les utilisateurs à utiliser cet URL lorsqu'ils souhaitent faire référence à une copie numérique.

Utilisation

4. Gratuité

Les EUB et les Bibliothèques de l'ULB mettent **gratuitement** à la disposition du public les copies numériques d'œuvres littéraires dont les exemplaires imprimés sont épuisés : aucune rémunération ne peut être réclamée par des tiers ni pour leur consultation, ni au prétexte du droit d'auteur.

5. Buts poursuivis

Les copies numériques peuvent être utilisées à des fins de recherche, d'enseignement ou à usage privé. Quiconque souhaitant utiliser les copies numériques à d'autres fins et/ou les distribuer contre rémunération est tenu d'en demander l'autorisation aux EUB, en joignant à sa requête, l'auteur, le titre, et l'éditeur du (ou des) document(s) concerné(s).

Demande à adresser aux Editions de l'Université de Bruxelles (EDITIONS@admin.ulb.ac.be).

6. Citation

Pour toutes les utilisations autorisées, l'utilisateur s'engage à citer dans son travail, les documents utilisés, par la mention « Université Libre de Bruxelles – Editions de l'Université de Bruxelles et Bibliothèques » accompagnée des précisions indispensables à l'identification des documents (auteur, titre, date et lieu d'édition).

7. Liens profonds

Les liens profonds, donnant directement accès à une copie numérique particulière, sont autorisés si les conditions suivantes sont respectées :

1. les sites pointant vers ces documents doivent clairement informer leurs utilisateurs qu'ils y ont accès via le site web des bibliothèques de l'ULB ;
2. l'utilisateur, cliquant un de ces liens profonds, devra voir le document s'ouvrir dans une nouvelle fenêtre ; cette action pourra être accompagnée de l'avertissement 'Vous accédez à un document du site web des bibliothèques de l'ULB'.

Reproduction

8. Sous format électronique

Pour toutes les [utilisations autorisées](#) mentionnées dans ce règlement le téléchargement, la copie et le stockage des copies numériques sont permis ; à l'exception du dépôt dans une autre base de données, qui est interdit.

9. Sur support papier

Pour toutes les [utilisations autorisées](#) mentionnées dans ce règlement les fac-similés exacts, les impressions et les photocopies, ainsi que le copié/collé (lorsque le document est au format texte) sont permis.

10. Références

Quel que soit le support de reproduction, la suppression des références aux EUB et aux Bibliothèques de l'ULB dans les copies numériques est interdite.